## PCT

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



#### INTERNATIONAL APPLICA

#### N PUBLISHED UNDER THE PATENT O

ERATION TREATY (PCT)

(51) International Patent Classification 0:

C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

A1

(11) International Publication Number:

WO 98/07864

(43) International Publication Date:

26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4 9625996.5 23 August 1996 (23.08.96) . G

13 December 1996 (13.12.96)

) GB

(60) Parent Application or Grant

(63) Related by Continuation

US Filed on 08/782,893 (CIP) 27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CENTRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford,
Charles [GB/GB]; Microbiological Research Authority
CAMR (Centre for Applied Microbiology & Research),
Porton Down, Salisbury, Wiltshire SP4 0JG (GB), QUINN,

Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

Al.	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	1_uxembourg	SN	Senegal
AU	Australia	GA	Gabon	1.V	1.atvia	<b>S7.</b>	Swaziland
A7.	Azerbaijan	GB	United Kingdom	MC	Мопасо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	1L	Israel	MR	Mauritania	· UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
Cl	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

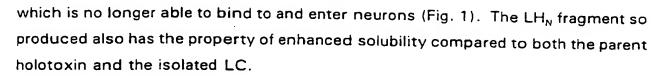
#### RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the  $H_{\rm C}$  domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the  $H_{\rm N}$  domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The  $H_{\rm N}$  domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the  $H_{\rm C}$ , amino acid residues 449-871 for the  $H_{\rm N}$  and residues 1-448 for the LC. Digestion with trypsin effectively degrades the  $H_{\rm C}$  domain of the BoNT/A to generate a non-toxic fragment designated  $LH_{\rm N}$ ,



It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

### (A) clostridial neurotoxin light chain:

-a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma - membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.

## (B) clostridial neurotoxin heavy chain H<sub>N</sub> domain:

- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H<sub>c</sub> domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H<sub>N</sub> of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H<sub>C</sub> domain.

For the purposes of the invention, the functional property or properties of the  $H_N$  of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a  $H_N$  domain or to the functions of a  $H_N$  domain are references to this property or properties. The second domain is not required to exhibit other properties of the  $H_N$  domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH<sub>423</sub>/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and  $H_N$  components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain  $H_N$  portion or a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the  $H_N$  domain. Teachings of regions within the  $H_N$  responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the  $H_N$  domain or fragment, though it too retains the function of the  $H_N$  domain. It is conveniently obtained by insertion, deletion and/or substitution of a  $H_N$  domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a  $H_N$  domain or fragment, (iii) a C-terminal extension to a  $H_N$  domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated  $H_N$  of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and  $H_N$  sequences of botulinum toxin types A, B, C<sub>1</sub>, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the  $H_{\rm C}$  domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated  $LH_{423}/A$  (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated  $LH_{728}/B$  (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The  $LH_N/A$  derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the  $H_C$  domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH<sub>423</sub>/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H<sub>C</sub> domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant  $LH_N$  derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH<sub>N</sub> enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

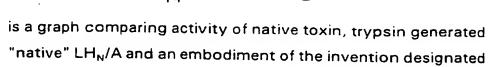
SEQ ID NO: 2, containing the entire L-chain and 423 LH423/A amino acids of the H-chain of botulinum neurotoxin type A; a variant of this molecule, containing a two amino acid 2LH423/A extension to the N-terminus of the L-chain; a further variant in which the molecule contains a two 2L12H423/A amino acid extension on the N-terminus of both the Lchain and the H-chain; <sub>2</sub>L<sub>FXa/2</sub>H<sub>423</sub>/A a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the Hchain component; and  $_{2}L_{FXa/2}H_{423}/A-IGF-1 =$ a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH<sub>423</sub>/A;

Fig. 3



 $_{2}LH_{423}/A$  ( $Q_{2}E,N_{26}K,A_{27}Y$ ) in an *in vitro* peptide cleavage assay;

- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/<sub>4</sub>H<sub>423</sub>/A illustrating insertion of four amino acids at the N-terminus of the H<sub>N</sub> sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H<sub>N</sub> sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated L<sub>FXa/3</sub>H<sub>423</sub>/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H<sub>N</sub> will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position  $G_{882}$ ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position  $\Omega_{882}$ ;
- Fig.9 shows the C-terminal portion of the amino acid sequence of an

embodiment of the invention designated  $L_{FXa/3}H_{423}/A-ZZ$ , a fusion protein; the C-terminal ZZ sequence begins at position  $A_{890}$  immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11

polypeptides of the invention; Fig. 10 shows LH<sub>423</sub>/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 1.2

shows the trypsin sensitive activation region of a polypeptide of the invention:

Fig. 13

shows Western blot analysis of recombinant LH<sub>107</sub>/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH<sub>107</sub>/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH<sub>107</sub>/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH<sub>423</sub>/A

2	· LH <sub>423</sub> /A
3	DNA coding for $_{23}LH_{423}/A$ ( $Q_2E,N_{26}K,A_{27}Y$ ), of which an N-terminal portion is shown in Fig. 4.
4	<sub>23</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
5	DNA coding for <sub>2</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y), of which an N-
	terminal portion is shown in Fig.4
6	$_{2}LH_{423}/A (Q_{2}E,N_{26}K,A_{27}Y)$
7	DNA coding for native BoNT/A according to Binz et al
8 /	native BoNT/A according to Binz et al
9	DNA coding for L <sub>/4</sub> H <sub>423</sub> /A
10	L <sub>/4</sub> H <sub>423</sub> /A
11	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A
12	$L_{FXa}/_{3}H_{423}/A$
13	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
14	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
15	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
16	$L_{FXa}/_3H_{423}/A$ -CtxA14
17	DNA coding for L <sub>FXa/3</sub> H <sub>423</sub> /A-ZZ
18	$L_{FXa/3}H_{423}/A-ZZ$
19	DNA coding for LH <sub>728</sub> /B
20	LH <sub>728</sub> /B
21	DNA coding for LH <sub>417</sub> /B
22	LH <sub>417</sub> /B
23	DNA coding for LH <sub>107</sub> /B
24	LH <sub>107</sub> /B
25	DNA coding for $LH_{423}/A$ ( $Q_2E,N_{26}K,A_{27}Y$ )
26	$LH_{423}/A (Q_2E, N_{26}K, A_{27}Y)$
27	DNA coding for LH <sub>417</sub> /B wherein the first 274 bases are

28

modified to have an *E.coli* codon bias

DNA coding for LH<sub>417</sub>/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

#### Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H<sub>c</sub>) of botulinum neurotoxin type A. This recombinant product is designated the LH<sub>423</sub>/A fragment (SEQ ID NO: 2).

#### Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the  $LH_{423}/A$  coding sequence was PCR amplified from total chromosomal DNA from  $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E. coli* plasmid vectors for expression analysis.

# Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth Expression of the gene is then induced by addition isopropylthio- $\beta$ -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH<sub>423</sub>/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST-  $LH_{423}/A$  is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant  $LH_{423}/A$  is recovered in the non-adsorbed material.

## Construct variants

A variant of the molecule,  $LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of  $LH_{423}/A$  producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

Two further variants of the gene sequence that have been expressed and the corresponding products purified are  $_{23}LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L<sub>/4</sub>H<sub>423</sub>/A (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH<sub>423</sub>/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L<sub>FXa/3</sub>H<sub>423</sub>/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L<sub>74</sub>H<sub>423</sub>/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of  $L_{FXa/3}H_{423}/A$  have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1)  $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2)  $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3)  $L_{\rm FXa/3}H_{423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

- 21 -

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing  $H_N$  function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the  $H_N$  domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

- 22 -

#### Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A ( $Q_2$ E, $N_{26}$ K, $A_{27}$ Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600  $\mu$ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10  $\mu$ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$ , 1.8 mM KH $_2$ PO $_4$ , pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

<sub>2</sub>LH<sub>423</sub>/A is more stable in the presence of trypsin and more active in the in vitro

peptide cleavage assay than is 23LH<sub>423</sub>/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

#### Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH<sub>728</sub>/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH<sub>417</sub>/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native  $\text{LH}_\text{N}/\text{A}$  .

A gene has also been assembled coding for a variant polypeptide,  $LH_{107}/B$  (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

#### **Construct Variants**

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

#### Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-<sub>2</sub>LH<sub>423</sub>/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH<sub>N</sub>/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na<sub>2</sub>HPO<sub>4</sub> 1.15 g/l, KH<sub>2</sub>PO<sub>4</sub> 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20  $\mu$ g of GST-<sub>2</sub>LH<sub>423</sub>/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to-its derivative LH<sub>N</sub>/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na<sub>2</sub>HPO<sub>4</sub> anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5  $\mu$ g (5X10<sup>-6</sup> g) to 5 picograms (5X10<sup>-12</sup> g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded



over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti-  $GST_{-2}LH_{423}/A$  antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH<sub>423</sub>/A antiserum.

#### **Botulinum Toxin/mouse**

Survivors On Day	0.5µg	0.005µg	0.0005 <b>µ</b> g	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	4	4	4	4	4	. 4
2	-	4	4	4	4	4	4
3	·	<b>. 4</b>	4	4	4	4	4
4		4	4	4	4	4	4

<u>TABLE 2.</u> Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

#### **Botulinum Toxin/mouse**

Survivors On Day	0.5 <b>µ</b> g	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	0 -	O	ο .	. 0	2	4
. 2	- ·	•	•	-	•	O <sub>1</sub>	4
3	-		•	-	•	•	4
4	•		•		•	•	4

## Example 4

Expression of recombinant LH<sub>107</sub>/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a  $LH_N$  of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide  $LH_{107}/B$  (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in  $E.\ coli\ BL21\ (DE3)$  (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for  $LH_{423}/A$ .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

#### SEQUENCE LISTING

#### (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
  - (B) STREET: 14 Kensington Square
  - (C) CITY: London
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): W8 5HH
  - (A) NAME: FOSTER; Keith Alan
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: QUINN; Conrad Padraig
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
    - (C) CITY: Salisbury
    - (D) STATE: Wiltshire
    - (E) COUNTRY: UK
    - (F) POSTAL CODE (ZIP): SP4 0JG
    - (A) NAME: SHONE; Clifford Charles
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk

  - (B) COMPUTER: IBM PC compatible (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2616 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)

- 3

#### (ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2616

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
								AAA Lys								144
								GGA Gly								192
								TAC Tyr								240
								AAG Lys								288
								CGT Arg 105								336
								AGT Ser								384
								GTG Val								432
								ATC Ile								480
								GGC								528
								TAC Tyr 185								576
								GAG Glu							TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720

				•			•			0 1	-					
CGC	GTG Val	TTC Phe	: AAG : Lys	GTT Val 245	Asņ	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	76
GAA Glu	GTA Val	AGC Ser	Phe 260	Glu	GAA Glu	CTG	CGC Arg	ACG Thr 265	Phe	GGT	GGC Gly	CAT	GAT Asp 270	Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT	GTG Val	913
GGT Gly 305	Thr	ACT	GCT Ala	TCA Ser	TTA Leu 310	Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 33.5	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
Asn	Phe	Val 355	AAG Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	Asn	1104
Phe	Asp 370	Lys	GCC Ala	Val	Phe	Lys 3 <b>7</b> 5	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr	1152
Thr 385	Ile	Tyr	GAT Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400	1200
Phe	Asn	Gly	-	Asn 405	Thr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
Lys	Asn	Phe	ACT Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
Gly	Ile	Ile 435	ACT	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
Ala	Leu 450	Asn	GAT Asp	Leu	Cys	11e 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe	1392
			GAA Glu													1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	TAA naa	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT Tyr	TAT Tyr	Leu	ACC · Thr 505	TTT Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	GAA Glu	CCT Pro	1536

									-	32	-					•
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT Tyr	GAG Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA 11e 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT	CAG Gln 750	TAT	AAT Asn	2256
Gln	Tyr	Thr 755	GAG Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp	2304
TTA Leu	AGT Ser 770	Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	Ser	ATA Ile	AAT	AAA Lys	GCT Ala 780	Met	ATT Ile	AAT Asn	ATA Ile	2352

1	AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
1	ATC [le	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
Į	SAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
C	AA Sln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT. Asp	2544
1	le	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
7					TAT Tyr			TAA			•		·				2616

#### (2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

150

- (A) LENGTH: 872 amino acids
- (B) TYPE: amino acid
  (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2: Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 55 Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys .120 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

170



Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Ğlu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295	Thr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val
Gly 305	Thr	Thr	Ala	Ser	Leu 310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	Asn
Phe	Asp 370	Lys	Ala	Val	Phe	Lys 375	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr
Thr 385	Ile	Tyr	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala		Asn 400
Phe	Asn	Gly	Gln	Asn 405	Thr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu
Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg
Gly	Ile	Ile 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys
Ala	Leu 450	Asn	Asp	Leu	Cys	Ile 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe
Ser 465		Ser	Glu	Asp	Asn 470	Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480
Ile	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490		Glu	Asn	Ile	Ser 495	Leu
Asp	Leu	Ile	Gln 500	Gln	Tyr	Tyr	Leu	Thr 505		Asn	Phe	Asp	Asn 510	Glu	Pro
Glu	Asn	Ile 515		Ile	Glu	Asn	Leu 520		Ser	Asp	Ile	Ile 525	Gly	Gln	Leu

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Cly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys . Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys

(2) INFORMATION FOR SEQ ID NO: 3:

WO 98/07864 PCT/GB97/02273 - 36 -

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2685 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1..2685
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	4.8	l
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	96	į
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	144	
AAG Lys	TAC Tyr 50	GGC Gly	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	AAG Lys 60	ATT Ile	CAT His	AAC Asn	AAA Lys	192	
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	240	I
GAC Asp	TTG Leu	AAC Asn	CCG Pro	CCG Pro 85	CCG Pro	GAA Glu	GCA Ala	AAG Lys	CAG Gln 90	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr 95	TAC Tyr	288	
GAT Asp	TCA Ser	ACC Thr	TAT Tyr 100	CTG Leu	AGC Ser	ACA Thr	GAC Asp	AAC Asn 105	GAG Glu	AAG Lys	GAT Asp	AAC Asn	TAC Tyr 110	CTG Leu	AAG Lys	336	
GGA Gly	GTG Val	ACC Thr 115	AAA Lys	TTA Leu	TTC Phe	GAG Glu	CGT Arg 120	ATT Ile	TAT Tyr	TCC Ser	ACT Thr	GAC Asp 125	CTG Leu	GGC Gly	CGT Arg	384	
ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	ATC Ile	GTC Val 135	CGC Arg	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser	432	
			ACG Thr													480	I
			GAC Asp										Leu			528	
			TCC Ser 180													576	
CAC His	GAA Glu	GTG Val 195	TTG Leu	AAC Asn	CTG Leu	ACG Thr	CGT Arg 200	AAC Asn	GGT Gly	TAC Tyr	GGC Gly	TCT Ser 205	ACT Thr	CAG Gln	TAC Tyr	624	:

						,	•										
					GAC Asp												672
					CTG Leu 230												720
GCG Ala	GTG Val	ACC Thr	CTG Leu	GCA Ala 245	CAC His	GAG Glu	CTG Leu	ATC Ile	CAC His 250	GCC Ala	GGT Gly	CAT His	CGT	CTG Leu 255	TAT Tyr		768
					CCG Pro												816
					GGT Gly												864
					GCG Ala												912
					TAC Tyr 310												960
					ATT												1008
					GAG Glu											•	1056
					AAA Lys												1104
					GAG Glu											•	1152
Asn	Arg	Lys	Thr	Tyr	TTG Leu 390	Asn	Phe	Asp	Lys	Ala	Val	Phe	Lys	Ile	Asn		1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg		1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC Asn	TTT Phe	AAT Asn 425	GGT Gly	CAA Gln	AAT Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn		1296
			Phe		AAA Lys												1344
Tyr					GTA Val												1392
					AAT Asn 470												1440
															-		

	-											•					
AAT Asn	'AAT Asn	TGG	GAC Asp	TTG Leu 485	Pne	TTT Phe	AGT Ser	CCT Pro	TCA Ser 490	Glu	GAT Asp	AAT Asn	TTT Phe	ACT Thr 495	AAT Asn	-	1488
GAT Asp	CTA Leu	AAT Asn	AAA Lys 500	GTÀ	GAA Glu	GAA Glu	ATT	ACA Thr 505	Ser	GAT Asp	ACT Thr	AAT Asn	ATA Ile 510	GAA Glu	GCA Ala		1536
GCA Ala	GAA Glu	GAA Glu 515	AAT Asn	ATT	AGT Ser	TTA Leu	GAT Asp 520	TTA Leu	ATA Ile	CAA Gln	CAA Gln	TAT Tyr 525	TAT	TTA Leu	ACC Thr		1584
TTT Phe	AAT Asn 530	TTT Phe	GAT Asp	AAT Asn	GAA Glu	CCT Pro 535	GAA Glu	AAT Asn	ATT Ile	TCA Ser	ATA Ile 540	GAA Glu	AAT Asn	CTT Leu	TCA Ser		1632
AGT Ser 545	GAC Asp	ATT Ile	ATA Ile	GGC	CAA Gln 550	TTA Leu	GAA Glu	CTT Leu	ATG Met	CCT Pro 555	AAT Asn	ATA Ile	GAA Glu	AGA Arg	TTT Phe 560		1680
CCT Pro	AAT Asn	GGA Gly	AAA Lys	AAG Lys 565	TAT Tyr	GAG Glu	TTA Leu	GAT Asp	AAA Lys 570	TAT Tyr	ACT Thr	ATG Met	TTC Phe	CAT His 575	TAT Tyr		1728
CTT Leu	CGT Arg	GCT Ala	CAA Gìn 580	GAA Glu	TTT Phe	GAA Glu	CAT His	GGT Gly 585	AAA Lys	TCT Ser	AGG Arg	ATT Ile	GCT Ala 590	TTA Leu	ACA Thr		1776
AAT Asn	TCT	GTT Val 595	AAC Asn	GAA Glu	GCA Ala	TTA Leu	TTA Leu 600	AAT Asn	CCT	AGT Ser	CGT Arg	GTT Val 605	TAT Tyr	ACA Thr	TTT Phe		1824
TTT Phe	TCT Ser 610	TCA Ser	GAC Asp	TAT Tyr	GTA Val	AAG Lys 615	AAA Lys	GTT Val	AAT Asn	AAA Lys	GCT Ala 620	ACG Thr	GAG Glu	GCA Ala	GCT Ala		1872
ATG Met 625	TTT Phe	TTA Leu	GGC Gly	TGG Trp	GTA Val 630	GAA Glu	CAA Gln	TTA Leu	GTA Val	TAT Tyr 635	GAT Asp	TTT Phe	ACC Thr	GAT Asp	GAA Glu 640		1920
ACT Thr	AGC Ser	GAA Glu	GTA Val	AGT Ser 645	ACT Thr	ACG Thr	GAT Asp	AAA Lys	ATT Ile 650	GCG Ala	GAT Asp	ATA Ile	ACT Thr	ATA Ile 655	ATT Ile		1968
ATT Ile	CCA Pro	TAT Tyr	ATA Ile 660	GGA Gly	CCT Pro	Ala	Leu	AAT Asn 665	Ile	Gly	Asn	Met	Leu	Tyr	AAA Lys		2016
GAT Asp	GAT Asp	TTT Phe 675	GTA Val	GGT Gly	GCT Ala	TTA Leu	ATA Ile 680	TTT Phe	TCA Ser	GGA Gly	GCT Ala	GTT Val 685	ATT. Ile	CTG Leu	TTA Leu		2064
GAA Glu	TTT Phe 690	ATA Ile	CCA Pro	GAG Glu	ATT Ile	GCA Ala 695	ATA Ile	CCT Pro	GTA Val	TTA Leu	GGT Gly 700	ACT Thr	TTT Phe	GCA Ala	CTT Leu		2112
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	GCG Ala	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720		2160
GCT Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg 725	AAT Asn	GAA Glu	AAA Lys	TGG Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile		2208
GTA Val	ACA Thr	AAT . Asn	TGG Trp 740	TTA Leu	GCA Ala	AAG Lys	GTT Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	GAT Asp	CTA Leu 750	ATA Ile	AGA Arg		2256

													•			
AAA Lys	AAA Lys	ATG Met 755	AAA Lys	GAA Glu	GCT	TTA Leu	GAA Glu 760	AAT Asn	CAA Gln	GCA Ala	GAA Glu	GCA Ala 765	ACA Thr	AAG Lys	GCT Ala	2304
ATA Ile	ATA Ile 770	AAC Asn	TAT	CAG Gln	TAT Tyr	AAT Asn 775	CAA Gln	TAT Tyr	ACT Thr	GAG Glu	GAA Glu 780	GAG Glu	AAA Lys	AAT Asn	AAT Asn	2352
ATT Ile 785	AAT Asn	TTT Phe	AAT Asn	ATT Ile	GAT Asp 790	GAT Asp	TTA Leu	AGT Ser	TCG Ser	AAA Lys 795	CTT Leu	AAT Asn	GAG Glu	TCT Ser	ATA Ile 800	2400
AAT Asn	AAA Lys	GCT Ala	ATG Met	ATT Ile 805	AAT Asn	ATA Ile	AAT Asn	AAA Lys	TTT Phe 810	TTG Leu	AAT Asn	CAA Gln	TGC Cys	TCT Ser 815	GTT Val	2448
TCA Ser	TAT Tyr	TTA Leu	ATG Met 820	AAT Asn	TCT Ser	ATG Met	ATC Ile	CCT Pro 825	TAT	GGT Gly	GTT Val	AAA Lys	CGG Arg 830	TTA Leu	GAA Glu	2496
GAT Asp	TTT Phe	GAT Asp 835	GCT Ala	AGT Ser	CTT Leu	AAA Lys	GAT Asp 840	GCA Ala	TTA Leu	TTA Leu	AAG Lys	TAT Tyr 845	ATA Ile	TAT Tyr	GAT Asp	2544
AAT Asn	AGA Arg 850	GGA Gly	ACT Thr	TTA Leu	ATT Ile	GGT Gly 855	CAA Gln	GTA Val	GAT Asp	AGA Arg	TTA Leu 860	AAA Lys	GAT Asp	AAA Lys	GTT Val	2592
AAT Asn 865	AAT Asn	ACA Thr	CTT Leu	AGT Ser	ACA Thr 870	GAT Asp	ATA Ile	CCT. Pro	TTT Phe	CAG Gln 875	CTT Leu	TCC Ser	AAA Lys	TAC Tyr	GTA Val 880	2640
									ACT Thr 890					TAA * 895		2685

### (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 895 amino acids
  - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20

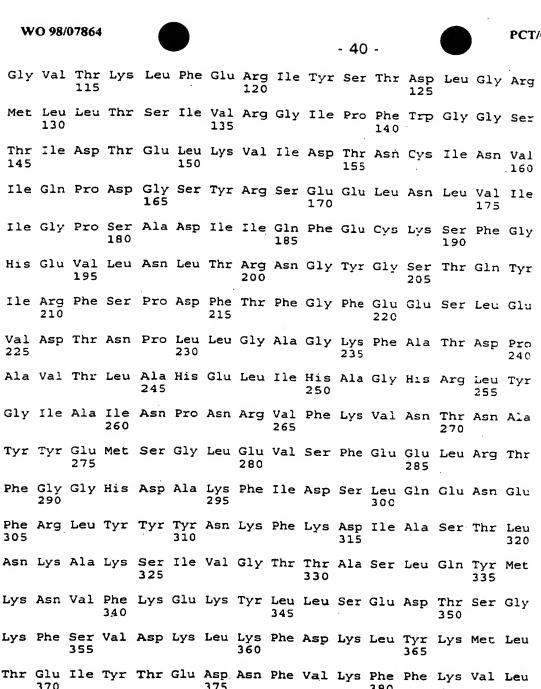
Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 70

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys



Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser

													_		
Leu 463	Asp	Lys	Gly	Tyr	Asn 470	Lys	Ala	Leu	Asn	Asp 475	Leu	Cys	Ile	Lys	Val 480
Asn	Asn	Trp	Asp	Leu 485	Phe	Phe	Ser	Pro	Ser 490	Glu	Asp	Asn	Phe	Thr 495	Asn
Asp	Leu	Asn	Lys 500		Glu	Glu	Ile	Thr 505	Ser	Asp	Thr	Asn	Ile 510	Glu	Ala
Ala	Glu	Glu 515	Asn	Ile	Ser	Leu	Asp 520	Leu	Ile	Gln	Gln	Tyr 525	Tyr	Leu	Thr
Phe	Asn 530	Phe	Asp	Asn	Glu	Pro 535		Asn	Ile	Ser	Ile 540	Glu	Asn	Leu	Ser
Ser 545	Asp	Ile	Ile	Gly	Gln 550	Leu	Glu	Leu	Met	Pro 555	Asn	Ile	Glu	Arg	Phe 560
Pro	Asn	Gly	Lys	Lys 565	Tyr	Glu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	Tyr
.Leu	Arg	Ala	Gln 580	Glu	Phe	Glu	His	Gly 585	Lys	Ser	Arg	Ile	Ala 590	Leu	Thr
Asn	Ser	Val - <b>5</b> 95	Asn	Glu	Ala		Leu 600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe
Phe	Ser 610	Ser	Asp	Tyr	Val	Lys 615	Lys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala
Met 625	Phe	Leu	Gly	Trp	Val 630	Glu	Gln	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640
Ţhr	Ser	Glu	Val	Ser 645	Thr	Thr	Asp	Lys	11e 650	Ala	Asp	Ile	Thr	11ė 655	Ile
Ile	Pro	Tyr	11e 660	Gļy	Pro	Ala	Leu	Asn 665	Ile	Gly	Asn	Met	Leu 670	Tyr	Lys
Asp	Asp	Phe 675	Val	Gly	Ala	Leu	Ile 680	Phe	Ser	Gly	Ala	Val 685	Ile	Leu	Leu
Glu	Phe 690	Ile.	Pro	Glu	Ile	Ala 695	Ile	Pro	Val	Leu	Gly 700	Thr	Phe	Ala	Leu
Val 705	Ser	Tyr	Ile	Ala	Asn 710	Lys	Val	Leu	Thr	Val 715	Gln	Thr	Ile	Asp	Asn 720
Ala	Leu	Ser	Lys	Arg 725	Asn	Glu	Lys	Trp	Asp 730	Glu	Val	Tyr	Lys	Tyr 735	Ile
Val	Thr	Asn	Trp 740	Leu	Ala	Lys	Val	Asn 745	Thr	Gln	Ile	Asp	Leu 750	Ile	Arg
Lys	Lys	Met 755	Lys	Glu · .	Ala	Leu	Glu 760.		Gln	Ala	Glu	Ala 765	Thr	Lys	Ala
Ile	Ile 770	Asn	Tyr	Gln	Tyr	Asn 775	Gln	Tyr	Thr	Glu	Glu 780	Glu	Lys	Asn	Asn
Ile 785	Asn	Phe	Asn	Ile	<b>Asp</b> 790	Asp	Leu	Ser	Ser	Lys 795	Leu	Asn	Glu	Ser	Ile 800
Asn	Lys	Ala	Met	Ile 805	Asņ	Ile	Asn	Lys	Phe 810	Leu	Asn	Gln	Cys	Ser 815	Val

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
865 Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys \*
895

# (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2622 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 1.. 2622
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA Gly 1	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT	AAG Lys	GAC Asp	CCT Pro 15	GTA Val		48
								AAA Lys 25									96
CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro		144
GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro		192
CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80		240
AGC Ser	ACA Thr	GAC Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	. •	288
TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser		336
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu		384
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly		432

						<b>'</b>											
AGC Ser 145	Tyr	AGA Arg	TCT	GAA Glu	GAA Glu 150	Leu	AAC Asn	CTC Leu	GTA Val	Ile 155	Ile	GGG Gly	CCC Pro	TCC Ser	GCG Ala 160		480
GAC Asp	ATT Ile	ATC Ile	CAG	TTT Phe 165	Glu	TGC Cys	AAG Lys	AGC Ser	TTT Phe 170	Gly	CAC His	GAA Glu	GTG Val	TTG Leu 175	AAC Asn		528
CTG Leu	ACG Thr	CGT Arg	AAC Asn 180	Gly	TAC Tyr	GGC Gly	TCT Ser	ACT Thr 185	CAG Gln	TAC Tyr	ATT Ile	CGT	TTC Phe 190	Ser	CCA Pro		576
GAC Asp	TTC Phe	ACG Thr 195	TTC Phe	GGT Gly	TTC Phe	GAG Glu	GAG Glu 200	AGC Ser	CTG Leu	GAG Glu	GTT Val	GAT Asp 205	ACC Thr	AAC Asn	CCG Pro		624
CTG Leu	TTG Leu 210	GGT	GCA Ala	GGC Gly	AAG Lys	TTC Phe 215	GCA Ala	ACT Thr	GAT Asp	CCA Pro	GCG Ala 220	GTG Val	ACC Thr	CTG Leu	GCA Ala	•	672
CAC His 225	GAG Glu	CTG Leu	ATC Ile	CAC His	GCC Ala 230	GGT Gly	CAT His	CGT Arg	CTG Leu	TAT Tyr 235	GGC	ATT	GCG Ala	ATT Ile	AAC Asn 240		720
CCG Pro	AAC Asn	CGC Arg	GTG Val	TTC Phe 245	AAG Lys	GTT Val	AAC Asn	ACC Thr	AAC Asn 250	GCC Ala	TAC Tyr	TAC Tyr	GAG Glu	ATG Met 255	AGT Ser		768
GGT Gly	TTA Leu	GAA Glu	GTA Val 260	AGC Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG Thr	TTC Phe	GGT Gly	GGC Gly 270	CAT His	GAT Asp	.*.	816
GCG Ala	AAG Lys	TTT Phe 275	ATC Ile	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	Phe	CGT Arg 285	CTG Leu	TAC Tyr	TAC		864
TAC Tyr	AAC Asn 290	AAG Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	AAG Lys	TCC Ser		912
									TAT Tyr								960
GAG Glu	AAA Lys	Tyr	Leu	Leu	Ser	Glu	Asp	Thr	TCT Ser 330	Gly	Lys	Phe	Ser	GTA Val 335	Asp		1008
									ATG Met								1056
									GTA Val								1104
									ATA Ile			Val					1152
									TTA Leu								1200
									ATT Ile 410								1248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	GAA Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC Tyr	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480	1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile	1488
AGT Ser	TTA Leu	GAT Asp	TTA Leu 500	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 505	TTA Leu	ACC Thr	TTT Phe	AAT Asn	TTT Phe 510	GAT Asp	AAT Asn	1536
GAA Glu	CCT Pro	GAA Glu 515	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 520	AAT Asn	CTT Leu	TCA Ser	AGT Ser	GAC Asp 525	ATT Ile	ATA Ile	GGC Gly	1584
CAA Gln	TTA Leu 530	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 535	ATA Ile	GAA Glu	AGA Arg	TTT Phe	CCT Pro 540	AAT Asn	GGA Gly	AAA Lys	AAG Lys	1632
TAT Tyr 545	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 550	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 555	CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 560	1680
TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 565	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 570	ACA Thr	AAT Asn	TCT Ser	GTT Val	AAC Asn 575	GAA Glu	1728
				CCT Pro												1776
				TAA neA												1824
				GTA Val												1872
				ATT Ile												1920
				ATA Ile 645												1968
				TCA Ser												2016
				GTA Val												2064

AAT Asn	AAG Lys 690	Val	CTA	ACC Thr	GTT Val	CAA Gln 695	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 700	TTA Leu	AGT Ser	AAA Lys	AGA Arg		.2112
AAT Asn 705	Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 710	GTC Val	TAT	AAA Lys	TAT Tyr	ATA Ile 715	GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 720		2160
GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 725	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 730	AGA Arg	AAA Lys	AAA Lys	ATG Met	AAA Lys 735	GAA Glu		2208
					GCA Ala												2256
					GAG Glu												2304
GAT Asp	GAT Asp 770	TTA Lėu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 775	AAT Asn	GAG Glu	TCT Ser	ATA Ile	AAT Asn 780	AAA Lys	GCT Ala	ATG Met	ATT		2352
AAT Asn 785	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 790	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 795	TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 800		2400
TCT	ATG Met	ATC Ile	CCT Pro	TAT Tyr 805	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 810	GAA Glu	GAT Asp	TTT Phe	GAT Asp	GCT Ala 815	Ser		2448
CTT Leu	AAA Lys	GAT Asp	GCA Ala 820	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 825	TAT Tyr	GAT Asp	AAT Asn	AGA Arg	GGA Gly 830	ACT Thr	TTA Leu	•	2496
					AGA Arg			Asp									2544
					CAG Gln												2592
					GAA Glu 870				TAA *								2622

#### (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 874 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly 135 Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro 185 Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 215 His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 265 Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 310 315 Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 395

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 4.05 Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr. Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu 455 Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu 550 Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp 595 Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly . 635 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 680 Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 690 Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln

PCT/GB97/02273

Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu 820 Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys \* 870 865 (2) INFORMATION FOR SEQ ID NO: 7: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2613 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double
(D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 1..2613 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7: ATG CCA TTT GTT AAT AAA CAA TTT AAT TAT AAA GAT CCT GTA AAT GGT Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 48 GTT GAT ATT GCT TAT ATA AAA ATT CCA AAT GCA GGA CAA ATG CAA CCA 96 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 GTA AAA GCT TTT AAA ATT CAT AAT AAA ATA TGG GTT ATT CCA GAA AGA 144 Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 GAT ACA TTT ACA AAT CCT GAA GAA GGA GAT TTA AAT CCA CCA CCA GAA 192 Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu GCA AAA CAA GTT CCA GTT TCA TAT TAT GAT TCA ACA TAT TTA AGT ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr GAT AAT GAA AAA GAT AAT TAT TTA AAG GGA GTT ACA AAA TTA TTT GAG 288 Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

	-					<b>'</b>										
AGA Arg	ATT Ile	TAT Tyr	TCA Ser 100	ACT Thr	GAT Asp	CTT Leu	GGA Gly	AGA Arg 105	ATG Met	TTG Leu	TTA Leu	ACA Thr	TCA Ser 110	Ile	GTA Val	336
AGG Arg	GGA Gly	ATA Ile 115	Pro	TTT	TGG Trp	GGT	GGA Gly 120	AGT Ser	ACA Thr	ATA	GAT Asp	ACA Thr 125	GAA Glu	TTA Leu	AAA Lys	384
GTT Val	ATT Ile 130	GAT Asp	ACT Thr	AAT Asn	TGT Cys	ATT Ile 135	AAT Asn	GTG Val	ATA Ile	CAA Gln	CCA Pro 140	GAT Asp	GGT Gly	AGT Ser	TAT Tyr	432
AGA Arg 145	TCA Ser	GAA Glu	GAA Glu	CTT	AAT Asn 150	CTA Leu	GTA Val	ATA Ile	ATA	GGA Gly 155	CCC Pro	TCA Ser	GCT Ala	GAT Asp	ATT Ile 160	480
ATA Ile	CAG Gln	TTT	GAA Glu	TGT Cys 165	AAA Lys	AGC Ser	TTT Phe	GGA Gly	CAT His 170	GAA Glu	GTT Val	TTG Leu	AAT Asn	CTT Leu 175	ACG Thr	528
CGA Arg	AAT Asn	GGT Gly	TAT Tyr 180	GGC	TCT	ACT Thr	CAA Gln	TAC Tyr 185	ATT Ile	AGA Arg	TTT Phe	AGC Ser	CCA Pro 190	GAT Asp	TTT Phe	576
ACA Thr	TTT Phe	GGT Gly 195	Phe	GAG Glu	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAA Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	GTA Val	ACA Thr 220	TTA Leu	GCA Ala	CAT His	GAA Glu	 672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT Ile	AAT Asn	CCA Pro	AAT Asn 240	 720
AGG Arg	GTT Val	TTT Phe	AAA Lys	GTA Val 245	AAT Asn	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	TAT Tyr	GAA Glu	ATG Met	AGT Ser	GGG Gly 255	TTA Leu	768
					GAA Glu											816
	Ile	Asp	Ser	Leu	CAG Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr			864
Lys	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	GTA Val	912
					TTA Leu 310											960
					GAT Asp											1008
					TAC Tyr											1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

	5	$\sim$	

TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ilė	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT	ACT Thr	AAA Lys 415	CTA Leu	1248
									TAT Tyr							1296
									TTA Leu							1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
									GAT Asp							1440
									GCA Ala 490							1488
									TTT Phe							1536
									AGT Ser							1584
									Pro							1632
									CTT Leu							1680
CAT His	GGT Gly	AAA Lys	TCT	AGG Arg 565	ATT	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 525	TTA Leu	1728
									TTT Phe							1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920

													•				
TTA Leu	AAT ASD	ATA Ile	GGT Gly	AAT Asn 645	. Met	TTA Leu	TAT	AAA Lys	GAT Asp 650	Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu		1968
ATA Ile	TTI Phe	TCA Ser	GGA Gly 660	Ala	GTT Val	ATT	CTG Leu	Leu 665	Glu	TTT	ATA Ile	CCA Pro	GAG Glu 670	Ile	GCA Ala		2016
ATA Ile	CCT Pro	GTA Val 675	Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	Leu	GTA Val	TCA Ser	TAT	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys		2064
GTT Val	CTA Leu 690	ACC	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu		2112
AAA Lys 705	Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG	TTA Leu	GCA Ala	AAG Lys 720		2160
GTT Val	AAT	ACA Thr	CAĠ Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu		2208
GAA Glu	AAT	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn		2256
CAA Gln	TAT	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Äsn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp		2304.
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT	AAT Asn	ATA Ile		2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT	ATG Met 800		2400
ATC Ile	CCT Pro	TAT Tyr	GGT	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys		2448
Asp	Ala	Leu	TTA Leu 820	Lys	Tyr	Ile	Tyr	<b>Asp 825</b>	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly		2496
Gln	Val	Asp 835	AGA Arg	Leu	Lys	Asp 	Lys 840	Val	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp		2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	Asn	CAA Gin 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser		2592
	Phe		GAA Glu		_			-		••						•	2613

# (2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 871 amino acids
    (B) TYPE: amino acid

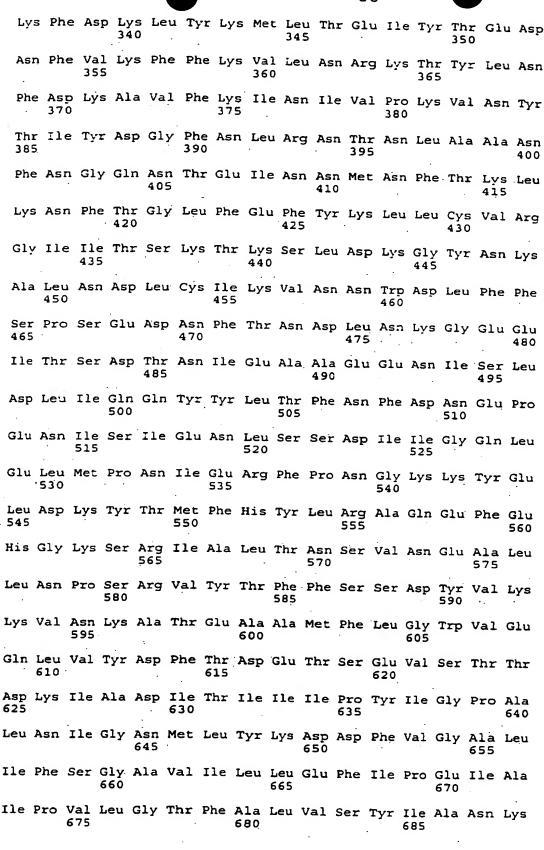
  - (D) TOPOLOGY: linear



- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu



PCT/GB97/02273

					,					- 54	-				
Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720
Val	Asn	Thr	Gln	Ile 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile
Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800
Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys
Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Туг	Asp 825	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 860

Thr Phe Thr Glu Tyr Ile Lys 865

## (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2628 base pairs

  - (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..2628
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

		AAG Lys						4.8
		ATC Ile						96
		ATT Ile						144
		CCG Pro						192

										•						
GCA Ala 65	Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	Asn	TAC	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT	TCC Ser 100	Thr	GAC Asp	CTG	GGC	CGT Arg 105	Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT	AAC Asn	CCG Pro	AAC Asn 240	720
CGC	GTG Val	TTC Phe	Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
										GGT Gly						816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
										AAG Lys						. 912
										AAT Asn 315						960
			Ser							TTT Phe						1008

PCT

AA/ Lys	A TTT	GA1	T AAG 5 Lys 340		A TAG	C AAJ	A ATO	349	Tur	A GAC	AT'	T TAG e Ty:	AC Th	r Gl	G GAT u Asp	1056
AAT Asr	TTTI Phe	GTI Val		TTT Phe	TTT	AA Lys	GTA Val 360	. ner	AA( Asr	AGA Arg	A AAI	A ACA	Ty	r TT	G AAT u Asn	1104
TTI Phe	GAT Asp 370	nya	GCC Ala	GTA Val	TTT Phe	Lys 375	TTE	AA1 : Asr	ATA	GTA Val	CCT Pro 380	Lys	GTA Val	A AA	TAC Tyr	1152
ACA Thr 385	110	TAT	GAT Asp	GGA Gly	TTT Phe 390	Wan	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	Ası	TTA Leu	GCA Ala	A GCA A Ala	A AAC A Asn 400	1200
FIIC	ASII	GIY	GIII	405	1111	GIU	ııe	Asn	410	Met	Asn	Phe	Thr	415		1248
Lys	ASII	FILE	420	GIY	Leu	Pne	GIU	425	Tyr	Lys	Leu	Leu	Cys 430	Val	AGA Arg	1296
Gly	116	435	1111	Ser	гус	inr	440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	AAG Lys	1344
261	450	Asp	GIY	ALA	Leu	455	Asp	Leu	Cys	Ile	Lys 460	Val	Asn	Asn	TGG	1392
465	Deu	FIIE	Pne	Ser	470	ser	GIU	Asp	Asn	Phe 475	Thr	Asn	Asp	Leu	AAT Asn 480	1440
дуз	GIY	Gru	Giu	485	1111	ser	Asp	Thr	490	Ile	Glu	Ala	Ala	Glu 495		1488
ASII	116	Sei	500	Asp	Leu	ııe	GIN	505	Tyr	TAT Tyr	Leu	Thr	Phe 510	Asn	Phe	1536
дар	ASII	515	PIO	GIU	ASI	iie	520	Tie	Glu	AAT Asn	Leu	Ser 525	Ser	Asp	Ile	1584
116	530	GIII	Leu	GIU	Leu	535	Pro	Asn	IIe	GAA Glu	Arg 540	Phe	Pro	Asn	Gly	1632
AAA Lys 545	nys	TYL	GIU	Leu	<b>Asp 550</b>	Lys	Tyr	Thr	Met	Phe 555	His	Tyr	Leu	Arg	Ala 560	1680
CAA Gln	GIU	File	GIU	565	GIŸ	Lys	ser	Arg	11e 570	Ala	Leu	Thr	Asn	Ser 575	Val	1728
AAC Asn	GIU	АТА	580	Leu .	Asn	PIO	ser .	Arg 585	Val	Tyr	Thr	Phe	Phe 590	Ser	Ser	1776
GAC Asp	TAL	GTA . Val : 595	AAG . Lys	AAA ( Lys	GTT . Val .	ASD .	AAA ( Lys ) 600	GCT Ala	ACG (	GAG ( Glu ,	GCA Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

											•					
G)	C TG y Tr 61	p va	A GA	A CAI u Gli	A TTA	Val 615	Tyr	GA:	TTT Phe	ACC Thr	GAT Asp 620	Glu	ACT Thi	AG(	GAA Glu	1872
GT Va 62	1 Se	T AC	T AC	G GAT	AAA Lys 630	: ile	GCG Ala	GAT Asp	T ATA	ACT Thr 635	Ile	ATT	ATT Ile	CCA Pro	TAT Tyr 640	1920
AT Il	A GG e Gl	A CC y Pr	T GCT o Ala	TTA Lev 645	AAT ASD	ATA Ile	GGT Gly	AA1	Met 650	Leu	TAT	AAA Lys	GAT Asp	GAT Asp 655	Phe	1968
GT. Va	A GGT	r gc y al	T TTA a Leu 660	1 116	·TTT Phe	TCA Ser	GGA Gly	GCT Ala 665	Val	ATT Ile	CTG Leu	TTA Leu	GAA Glu 670	Phe	ATA Ile	2016
CC. Pr	A GAG	3 AT 1 Ile 67	e ATS	ATA Ile	CCT Pro	GTA Val	TTA Leu 680	GGT Gly	ACT Thr	TTT Phe	GCA Ala	CTT Leu 685	GTA Val	TCA Ser	TAT Tyr	 2064
AT	T GCC e Ala 690	a Ası	T AAC n Lys	GTT Val	CTA Leu	ACC Thr 695	GTT Val	CAA Gln	ACA Thr	ATA Ile	GAT Asp 700	AAT Asn	GCT Ala	TTA Leu	AGT Ser	2112
AA Ly: 70:	s Arg	AA? ASI	r GAA n Glu	AAA Lys	TGG Trp 710	Asp	GAG Glu	GTC Val	TAT Tyr	AAA Lys 715	Tyr	ATA Ile	GTA Val	ACA Thr	AAT Asn 720	2160
TG( Trị	TTA Leu	GCA Ala	A AAG A Lys	GTT Val 725	AAT Asn	ACA Thr	CAG Gln	ATT Ile	GAT Asp 730	CTA Leu	ATA Ile	AGA Arg	AAA Lys	AAA Lys 735	ATG Met	2208
Lys	. Glu	Ala	740	GIu	AAT Asn	Gln	Ala	Glu 745	Ala	Thr	Lys	Ala	11e 750	Ile	Asn	2256
Tyr	Gin	755	Asn	Gin	TAT	Thr	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe	2304
Asn	770	Asp	Asp	Leu	AGT Ser	5er 775	Lys	Leu	Asn	Glu	<b>Ser</b> 780	Ile	Asn	Lys	Ala	2352
Met 785	Ile	Asn	Ile	Asn	AAA Lys 790	Phe	Leu	Asn	Gln	Cys 795	Ser	Val	Ser	Tyr	<b>Leu</b> 800	2400
Met	Asn	Ser	Met	11e 805	Pro	Tyr (	Gly	Val	B10	Arg	Leu	Glu	Asp	Phe 815	Asp	2448
Ala	Ser	Leu	<b>Lys</b> 820	Asp	GCA Ala	Leu 1	Leu .	Lys 825	Tyr	Ile	Tyr	Asp .	Asn 830	Arg .	Gly	2496
Thr	Leu	11e 835	Gly	Gln	GTA (	Asp 1	Arg :	Leu	Lys .	Asp :	rys	Val 1 845	Asn .	Asn	Thr	2544
Leu	Ser 850	Thr	Asp	Ile		Phe (	3ln 1	Leu	Ser 1	Lys '	1yr '	GTA ( Val i	GAT . Asp .	AAT Asn	CAA Gln	2592
AGA Arg 865	TTA Leu	TTA Leu	TCT Ser	Thr	TTT / Phe 7 870	ACT (	GAA :	rat . ryr	Ile 1	AAG : Lys 375	raa *					2628

#### (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 876 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val

300

295

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 315 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 470 Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 490 Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala - 550 Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650



Val	Gly	Ala	Leu 660	Ile	Phe	Ser	Gly	Ala 665	Val	Ile	Leu	Leu	Glu 670	Phe	Il
Pro	Glu	11e 675	Ala	Ile	Pro	Val	Leu 680	Gly	Thr	Phe	Ala	Leu 685	Val	Ser	Ty
Ile	Ala 690	Asn	Lys	Val	Leu	Thr 695	Val	Gln	Thr	Ile	Asp 700	Asn	Ala	Leu	Se
Lys 705	Arg	Asn	Glu	Lys	Trp 710	Asp	Glu	Val	Tyr	Lys 715	Tyr	Ile	Val	Thr	As:
Trp	Leu	Ala	Lys	Val 725	Asn	Thr	Gln	Ile	Asp 730	Leu	Ile	Arg	Lys	Lys 735	Met
Lys	Glu	Ala	Leu 740	Glu	Asn	Gln	Ala	Glu 745	Ala	Thr	Lys	Ala	Ile 750	Ile	Asr
Tyr	Gln	Tyr 755	Asn	Gln	Tyr	Thr	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe
Asn	Ile 770	Asp	Asp	Leu	Ser	Ser 775	Lys	Leu	Asn	Glu	Ser 780	Ile	Asn	Lys	Ala
Met 785	Ile	Asn	Ile	Asn	Lys 790	Phe	Leu	Asn.	Gln	Cys 795	Ser	Val	Ser	Tyr	Leu 800
Met	Asn	Ser	Met	Ile 805	Pro	Tyr	Gly	Val	Lys 810	Arg	Leu	Glu	Asp	Phe 815	Asp
Ala	Ser	Leu	Lys 820	Asp	Ala	Leu	Leu	Lys 825	Tyr	Ile	Tyr	Asp	Asn 830	Arg	Gly
Thr	Leu	Ile 835	Gly	Gln	Val	Asp	Arg 840	Leu	Lys	Asp	Lys	Val 845	Asn	Asn	Thr
Leu	Ser 850	Thr	Asp	Ile	Pro	Phe 855	Gln	Leu	Ser	Lys	Tyr 860	Val	Asp	Asn	Gln

# (2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2637 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys

- (ii) MOLECULE\_TYPE: DNA (genomic)
- (ix) FEATURE:

  - (A) NAME/KEY: CDS
    (B) LOCATION:1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT 48 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

875

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20

						,							•				
GT(	AAC Lys	G GC' S Ala	a Phe	C. AAC E Lys	S AT	Γ CAT ∋ His	AAC Asi 40	ı Lys	ATC Ile	Tr	GTT Val	T ATT	Pro	G GA	A CGC ù Arg		144
GAT Asp	ACA Thr	Phe	r AÇC e Thi	AAC ASI	CCC Pro	G GAA O Glu 55	Gli	A GGA 2 Gly	GAC Asp	TTC Leu	AAC Asr 60	Pro	CCC Pro	G CC	G GAA D Glu		192
GCA Ala 65	Lys	Glr	G GTC	CCA Pro	GTT Val	. Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thr	TAT	CTC Lev	G AGO	C ACA Thr 80		240
GAC Asp	AAC Asn	GAC Glu	AAC Lys	GAT Asp 85	Asr	TAC Tyr	CTC Leu	AAG Lys	GGA Gly 90	' Val	ACC Thr	Lys	TTA Leu	TTC Phe 95	GAG Glu		288
CGT	ATI	TAI	TCC Ser 100	Thr	GAC Asp	CTG Leu	GGC	CGT Arg 105	Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	: Ile	GTC Val		336
CGC	GGA Gly	ATC Ile	Pro	TTT Phe	TGG Trp	GGT	GGC Gly 120	Ser	ACC Thr	ATT	GAC Asp	ACG Thr	Glu	TTO Leu	AAG Lys		384
GTT Val	ATT Ile 130	Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	Asp	GGT Gly	AGC Ser	TAC		432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr		528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe		576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu		624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	Asp	Pro	GCG Ala	Val	Thr	CTG Leu	GCA Ala	CAC His	GAG Glu		672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240		720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	•	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys		816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn		864
Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	Ala	AGT Ser 295	ACA · Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val		912

- 62 -

GGT Gly 305	Thr	ACT Thr	GCI Ala	TCA Ser	TTA Leu 310	Gln	TAT	ATG Met	AAA Lys	AAT Asn 315	Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	Asp	ACA Thr	TCT Ser	GGA Gly	Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	AAC Asn	Arg	Lys	Thr 365	Tyr	Leu	Asn	1104
TTT Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	TIE	TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	TAA Asn	GAA Glu	CCT Pro	GAA Glu 520	AAT Asn	ATT	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
CCT Pro 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680
CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728

AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TA AST. Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Pleaser Ser Asp Tyr Val Lys Val Asn Lys Ala Thr Glu Ala Al Ser Ser Asp Tyr Val Lys Val Asn Lys Ala Thr Glu Ala Al Ser TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAMET Phe Leu Gly Typ Val Glu Gla Leu Val Typ Val Gly Cla Leu Val Typ Val Gly Gly Cla Leu Val Typ Val Gly Gly Cla Leu Val Typ Val Gly Gly Gly Cla Leu Val Typ Val Gly Gly Gly Gly Gly Gly Gly Gly Gly Gl	T 1824 la 1872
Prie Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Al 595 600 605	NA 1872
ATG TIT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GA	iu
Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Gl 610 615 620	T 1920
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA AT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Il 625 630 635	e
ATT CCA TAT ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AA Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Ly 645 650 655	A 1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TT ASP Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Le 660 665	A 2016 u
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CT Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Le 675 680 685	T 2064 u
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AA' Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Ass 690 695 700	T 2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715	•
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730 735	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn 755 760 765	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 770 775 780	2352
AAT AAA GCT ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800	
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815	2448
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830	2496
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845	2544

- 64 -

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA 2592 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA 2637 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870 875

- (2) INFORMATION FOR SEQ ID NO: 12:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 879 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12: Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 105 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 155 Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 205 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu

250

255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600

- 66 -

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

- (2) INFORMATION FOR SEQ ID NO: 13:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2862 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION: 1..2862
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

•																	
AT Me	G CA t Gl 1	G Ti	rc gr ne Va	TG A	AC A sn L 5	AG C. YS G.	AG TT	TC AA ne As	n Ty	AT ÁA Yr Ly .0	.G GA 's As	C CC	T G1 O Va	l As	AC GGT sn Gly	4	18
GT Va	T GA l As	C AT	E A.	CC TI la Ty 20	AC A'	TC AM	AA A7 /s I1	e Pr	A AA o As 5	C GC	Ç GG a Gl	C CA y Gl	n Me	G CA t Gl	G CCG	ġ	96
GT( Va.	G AA	2 MI	T TT a Ph	rc Al ne Ly	AG A? /s I]	TT CA	S AS	C AA n Ly 0	A AT s Il	C TG	G - GT p - Va	T AT 1 11 4	e Pr	G GA o Gl	A CGC u Arg	14	4
AS	5(	5	.6 11	II AS	ii Pi	5	5 GI	u GI	y As	p Le	AS:	n Pr	o Pr	o Pr	G GAA o Glu		2
65	L Lys		ıı va	ı Pi	5 va	0	r ly	r Ty	r As <sub>l</sub>	p Sei 75	r · Thi	r Ty:	r Le	u Se	C ACA r Thr 80	24	0
. vař	) MSI	. Gi	u Ly	8 8	p As 5	піу	r re	r ras	9 Gly	y Val	l Thi	r Lys	s Le	2 Ph		. 28	8
ALG		: Ту	10	0	I AS	b re	n GI	105	Met	Leu	. Leu	ı Thı	: Sei	r Ile	C GTC E Val	33	6
ALG	GIY	115	. PIC	o Pn	e ir	b GI	120	/ Ser	Thr	lle	Asp	125	Glu	ı Leı	AAG Lys	384	4
۷ĢΙ	130	vəf		. ASI	ı Cy:	135	AST	ı vaı	ile.	Gln	Pro 140	Asp	Gly	Ser	TAC	432	2
145	Ser	GI.	GIU	ı Let	150	Let	ı vaı	ile	Ile	GGG Gly 155	Pro	Ser	Ala	Asp	11e 160	480	)
	GIII	Pne	GIU	165	Lys	Ser	Pne	Gly	His 170	GAA Glu	Val	Leu	Asn	Leu 175	Thr	528	3
Arg	ASII	GIY	180	GIY	Ser	Inr	GIn	Tyr 185	Ile		Phe	Ser	Pro 190	Asp	Phe	576	
Titt	FIIE	195	Pne	GIU	ĠТU	Ser	200	GLu	Val	GAT Asp	Thr	Asn 205	Pro	Leu	Leu	624	
GIY	210	GIY	rys	Pne	Ala	215	Asp	Pro	Ala	GTG Val	Thr 220	Leu	Ala	His	Glu	672	
225	īřē	nis	AIA	GIY	230	Arg	Leu	Tyr	Gly.	ATT Ile 235	Ala	Ile,	Asn	Pro	Asn 240	720	
Arg	vai	rne	rys	245	Asn	Thr	Asn	Ala	Ту <del>г</del> 250	TAC Tyr	Glu	Met	Ser	Gly 255	Leu	768	
GAA ( Glu	GTA Z	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT (	GGC Gly	His	GAT Asp 270	GCG Ala	AAG Lys	816	

- 68 -

TTT Phe	ATC	GAC Asp 275	sex	TTC Lev	G CAG	GAG Glu	AAC Asn 280	Glu	TTC Phe	CGT	CTG Leu	TAC Tyr 285	TAC	TAC	AAC Asn	864
AAG Lys	Phe 290	rys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	Inr	ACT Thr	GCT Ala	TCA Ser	Leu 310	Gin	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT	GAT Asp	AAG Lys 340	Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Pḥe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	GAT Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

- 69 -

	CCT Pro 545	Asr	GG.	À AÀ y Ly	A AA s Ly	G TA S Ty 55	T GAG r Glu	TTA Leu	GA1	AAA Lys	TA1 5 Ty1 555	Thi	T ATO	G TTO	C CA' e Hi	TAT S Tyr 560	1680	
	CTT Leu	CGT Arg	GC Al	T CA a Gl	A GA n Gl	u Ph	T GAA e Glu	CAT His	GG1	AAA Lýs 570	Ser	AGC Arg	G ATT	GC:	TTI Let 579	A ACA 1 Thr	1728	
	AAT Asn	TCT Ser	G <b>T</b>	T AA l As 58	n Gli	A GCA u Ala	A TTA a Leu	TTA Leu	AAT Asn 585	Pro	AGT Ser	CGT Arg	GTT Val	TAT Ty:	Thi	A TTT	1776	
}	TTT Phe	TCT	TC Se 59	r As	C TAT	r GTA	A AAG L Lys	AAA Lys 600	Val	AAT Asn	' AAA Lys	GCT Ala	ACC Thr	: Glu	G GCA 1 Ala	A GCT A Ala	1824	
	ATG Met	TTT Phe 610	Lei	A GGC	C TGO y Tri	GTA Val	GAA Glu 615	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 620	Phe	ACC Thr	GAT Asp	GAA Glu	1872	
7	ACT Thr 525	AGC Ser	GA7 Glu	A GTA	A AG1 I Sei	T ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT	GCG Ala 635	Asp	ATA Ile	ACT Thr	T ATA	ATT Ile 640	1920	
3	ATT [le	CCA Pro	TAT	T ATA	GGA Gly 645	Pro	GCT Ala	TTA	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	Leu	TAT Tyr 655	Lys	1968	
P	SAT	GAT Asp	TTT Phe	GTA Val	Gly	GCŢ	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT	GTT Val	ATT Ile 670	Leu	TTA Leu	2016	
G	AA lu	TTT Phe	ATA Ile 675	Pro	GAG Glu	ATT	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	2064	
V	TA 'al	TCA Ser 690	TAT	ATT	GCG Ala	AAŤ Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA	GAT Asp	AAT Asn	2112	
A	CT la 05	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT	AAA Lys	TAT Tyr	ATA Ile 720	2160	
G V	TA al	ACA Thr	AAT Asn	TCC	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	 2208	
A.	AA ys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256	
A'	TA . le	ATA. Ile	AAC Asn 755	TAT	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn	2304	
A:	le A	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA . Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352	
As	AT A	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA . Ile .	AAT . Asn :	AAA Lys	Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	 2400	
T( Se	CA 1	rat fyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG A	ATC (	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448	

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	2688
TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	2736
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	2784
TGC Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	GAG Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	2832
				AAG Lys					TAG *			٠				2862

#### (2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 954 amino acids(B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

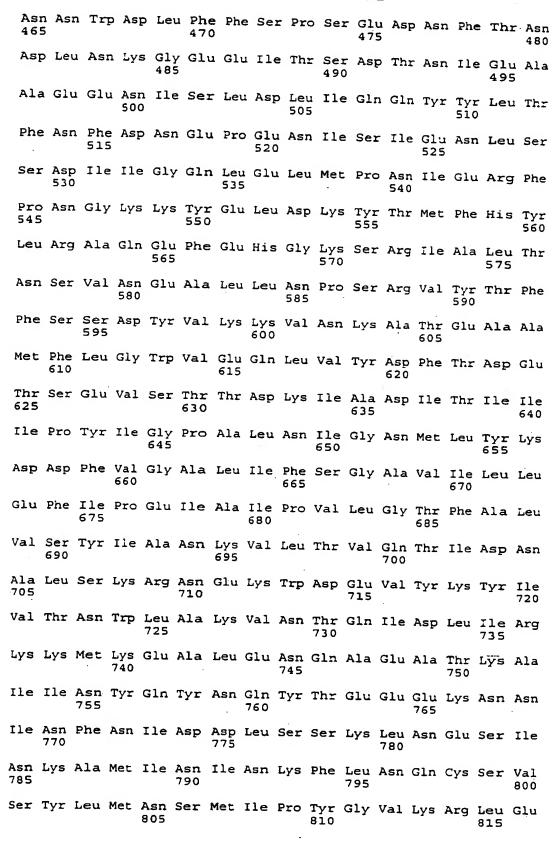
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 105

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Ard Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 455 460



						•					•					
As	p Ph	e As	p Ala 820	s Sei	r Lei	ı Lys	s Asp	829	a Lev	ı Lei	Ly:	з Ту	r Il		r Asp	
As	n Ar	g Gl: 83	y Thi	. Leu	ı Ile	e Gly	/ Glr 840	Val	Asp	Arg	g Let	Ly:		p Lys	s Val	
As	n Ası 850	n Th:	r Lei	ı Ser	Thr	855	Ile	Pro	Phe	e Glr	1 Let 860		r Ļys	з Туз	r Val	
As:	p Ası 5	n Gli	n Arg	, Lev	Leu 870	Ser	Thr	Phe	Thr	Glu 875		: Ile	⊇ Lys	s Ser	Arg 880	
Pro	o Gly	/ Pro	Glu	885	Leu	Cys	Gly	Ala	Glu 890	Leu	ı Val	. Asp	Ala	895	ı Gln	•
Pho	e Val	L Cys	900	Asp	Arg	Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	910		' Tyr	
Gly	y Ser	915	Ser	Arg	Arg	Ala	Pro 920	Gln	Thr	Gly	lle	Val 925		Glu	Cys	
Cys	Phe 930	Arg	Ser	Cys	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940		Cys	Ala	Pro	
Le: 945	_	Pro	Ala	. Lys	Ser 950	Ala	Glu	Ala	* •							
(2)	INF	ORMA	MOIT	FOR	SEQ	ID 1	NO:	15:								
	(i	) SE	QUEN	CE C	HARA	CTER	ISTI	CS:								
	,	(	A) L	ENGT	H: 2'	724 l	oase	pai	rs							
						leic ESS:										
						line									•	
	(ii	) MO	LECU	LE T	YPE:	DNA	(ger	nomi	=)						,	
			ATUR				_							٠.		
	(17	(.	A) N	AME/I						·						
		()	B) LO	DCAT]	ON: 1	127	724								•	•
	1		011E316	D-	cant											
						PTIC			•							
ATG	CAG	TTC	GTG	AAC	AAG	CAG Gln	TTC	AAC	TAT	AAG	GAC	CCT	GTA	AAC	GGT	. 48
1	GIII	-116	Val	5	цуз	·	FILE	ASII	10	гÀг	Asp	Pro	vai		GIA	
GTT	GAC	ATT	GCC	TÁC	ATC	AAA	ATT	CCA	246	GCC	GGC	CNG	איני	ĊNC	ccc	0.0
Val	Asp	Ile	Ala	Tyr	Ile	Lys	Ile	Pro	Asn	Ala	Gly	Gln	Met	Gln	Pro	96
	•		20					25		٠.			30			
GTG Val	AAG Lvs	GCT	TTC	AAG Lvs	ATT	CAT His	AAC	AAA	ATC	TGG	GTT	ATT	CCG	GAA	CGC	144
	-1-	35		-1-			40	2,3			Val.	45	FIO	GIU	Arg	
GAT	ACA	TTT	ACG	AAC	CCG	GAA (	GAA	GGÀ	GAC	TTG	AAC	CCG	CCG.	CCG	GAA	192
Asp	Thr 50	Phe	Thr	Asn	Pro	Glu ( 55	Glu	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	Glu	172
GCA:	AAG	CAG	GTG	CCA	GTT	TCA '	TAC	TAC	GAT	TCA	ACC	TAT	CTG	AGC	ACA	240
Ala 65	Lys	Gln	Val	Pro	Val . 70	Ser '	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr	
										_	•				80	•
GAC Asp	AAC Asn	GAG Glu	AAG (	GAT A	AAC ' Asn '	TAC (	CTG . Leu	AAG Lvs	GGA Glv	GTG Val	ACC Thr	AAA	TTA	TTC	GAG	288
			•	85		•		-, <b>-</b>	90			-, -		95	J. u	-

95

CG' Arg	T AT	TA:	r TCC r Sei		r Gad	CTC Lev	GGG Gly	C CGT / Arc	j Met	G CTC	CTC Lev	G ACC	TCA Ser	r Ile	C GTC ≥ Val		336
CGC	GG/ Gly	A ATO		TTT Phe	r TGC = Trp	GGT Gly	GG( Gly 120	ser	ACC Thr	ATI	GAC Asp	ACG Thr	Gli	TTC Let	AAG Lys		384
GT1 Val	116 130		ACT Thr	AAC Asr	TGC Cys	ATT Ile 135	ASI	GTG Val	ATC Ile	CAA Gln	CCA Pro	Asp	GGT Gly	AGC Ser	TAC		432
AGA Arc 145	, 361	GAA Glu	.GAA	CTI Lev	AAC Asn 150	Leu	GTA Val	ATC	ATC Ile	GGG Gly 155	Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
ATC	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	LYS	AGC Ser	TTT Phe	GGC	CAC His 170	Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr		528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GIY	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	٠.	576
ACG Thr	TTC Phe	GGT Gly 195	Pile	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu		624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu		672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240		720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu		768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys		816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn		864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val		912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320		960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu		1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp		1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn		1104

Į,

TT: Phe	GA Asj	b rà	A GC s Al	C GT	A TT:	r AAG E Lys	s Ile	A AA1 e Ası	T ATA	A-GTA e Val	A CCT	o Ly	G GT/ s Val	A AA' L Asi	TAC Tyr		1152
ACA Thi	r Ile	A TA e Ty	T GA' r Ası	r GG/ p Gly	A TTT / Phe 390	e Asr	TTA Lei	A AGA	A AAT J Asi	Thr 395	: Asi	T TT!	A GCA a Ala	A GCA A Ala	A AAC A Asn 400		1200
Phe	e Ası	n Gl	y Gl	1 Asr 409	i Thi	Glu	ı Ile	Asn	410	Met	: Ası	n Phe	Thr	415			1248
Lys	. Ası	ı Pho	e Thi 420	Gly	Leu	Phe	: Glu	Phe 425	Tyr	Lys	Leu	ı Lev	430	Val	AGA Arg		1296
Giy	· Ile	435	E Thr	Ser	. Lys	Thr	440	Ser	Leu	Asp	Lys	445	Туг	Asn	AAG Lys		1344
Ile	450	GI	/.Arg	, Cys	Asp	Gly 455	Ala	Leu	Asn	Asp	Leu 460	Cys	: Ile	Lys			1392
Asn 465	Asn	Trp	Asp	Leu	Phe 470	Phe	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	480		1440
Asp	· Leu	Asn	Lys	Gly 485	Glu	Glu	Ile	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495		•	1488
Ala	Glu	Ģlu	AAT Asn 500	Ile	Ser	Leu	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	Thr	٠	1536
Phe	Asn	Phe 515	GAT Asp	Asn	Glu	Pro	Glu 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Leu	Ser		1584
Ser	530	Ile	ATA Ile	Gly	Gln	Leu 535	Glu	Leu	Met	Pro	Asn 540	Ile	Glu	Arg	Phe		1632
Pro 545	Asn	Gly	AAA Lys	Lys	Tyr 550	Glu	Leu	Asp	Lys	Tyr 555	Thr	Met	Phe	His	Tyr 560		1680
Leu	Arg	Ala	CAA Gln	565	Phe	Glu	His	Gly	Lys 570	Ser	Arg	Ile	Ala	<b>Leu</b> 575	Thr		1728
Asn	Ser	Val	AAC Asn 580	Glu	Ala	Leu	Leu	Asn 585	Pro	Ser	Arg	Val	Tyr 590	Thr	Phe		1776
Phe	Ser	Ser 595	GAC Asp	Tyr	Val	Lys	Lys 600	Val	Asn	Lys	Ala	Thr 605	Glu	Ala	Ala		1824
Met	Phe 610	Leu	GGC	Trp	Val	Glu 615	Gln	Leu	Val	Tyr	Asp 620	Phe	Thr	Asp	Glu		1872
ACT Thr 625	AGC Ser.	GAA Glu	GTA Val	Ser	ACT ATTACT OF ACT OF AC	ACG (	GAT Asp	AAA . Lys	Ile .	GCG ( Ala / 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	٠	1920

															_	
AT Il	T CCA e Pro	TAI	I ATA	GGA Gly 645	, bro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	Lys	1968
GA As	T GAI p Asp	TTT Phe	GTA Val 660	GIV	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	Leu	TTA Leu	2016
GA Gl	A TTT u Phe	ATA	Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	2064
GT Va	A TCA l Ser 690	Tyr	ATT	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	Thr	ATA Ile	GAT Asp	AAT Asn	2112
GC Al 70	T TTA a Leu 5	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	2160
GT. Va	A ACA l Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
AA. Ly:	A AAA s Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256
AT	A ATA e Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn	2304
AT:	T AAT B Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
AAT Asi 789	r AAA n Lys 5	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400
TC# Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448
GA1 Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	Tyr	GAT Asp	2496
AA1 Asr	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	2544
AA1 Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	TAAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT	AGG Arg 880	2640
CCI	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA Gln	ATA Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	CAA Gln	TCT Ser 895	GAT Asp	2688
ATT	GAT Asp	ACA Thr	CAT His 900	AAT Asn	AGA Arg	ATT Ile	AAG Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA *					2724

#### (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 908 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

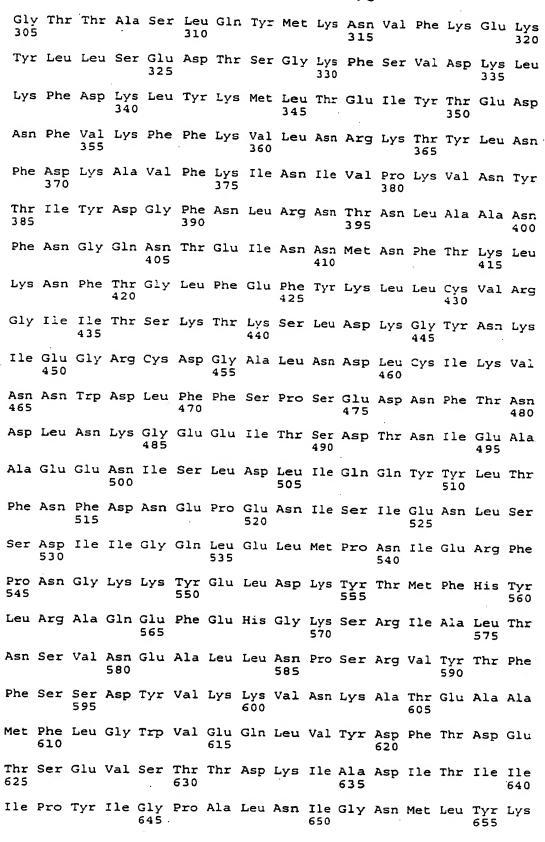
Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300



Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 790 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 890 Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu

### (2) INFORMATION FOR SEQ ID NO: 17:

900

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3042 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..3042
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

- 80 -

GTT Val	GAC Asp	ATT	GCC Ala 20	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT	AAC Asn	CCG Pro	AAC Asn 240	720
		TTC Phe														768
		AGC Ser														816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864

AA( Lys	G TT s Ph 29	е гу	VA GA	TA TA	T GC e Al	A AG a Se 29	r Th	A CTO	G AAG L Asi	C AAG	G GC" S Al.	a Ly:	G TC s Se	C AI r Il	T GTO	912
GG1 G13 305	y in	C AC	T GO	T TO	A TT r Le 31	n GT	G TA' n Ty:	T ATO	G AAJ	A AA1 5 Asr 315	ı Val	T TT	r aa. e Ly:	A GA s Gl	G AAA u Lys 320	•
TA1 Tyr	CT Le	C CT u Le	A TO u Se	T GA r Gl 32	u As	T AC	A TC	r GG/	A AAA / Lys 330	s Phe	C TCC	G GT/	A GA: L Ası	T AA P Ly 33	A TTA s Leu 5	1008
nys	; Pn	e AS	34	o Le	u Ty	r Ly:	s Met	345	i Thr	Glu	ı Ile	туг	350	r Gl	G GAT u Asp	
ASI	Pne	35:	1 Ly 5	s Pn	e Phe	i PAs	360	L Leu	Asn	Arg	Lys	365	Tyr	Le	G AAT u Asn	
Phe .	370	) )	S Al	a va	r Pne	375	; I1e	: Asn	Ile	· Val	9rc 380	Lys	Val	. Ası	TAC Tyr	1152
385	116	: ry:	r As	o Gr	390	AST	. Leu	Aṛg	Asn	395	Asn	. Leu	Ala	Ala	A AAC A Asn 400	1200
rne	Asn	GI	/ Glr	405	i Thr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		1248
Lys	Asn	Phe	.420	GIA	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	AGA Arg	1296
GIY	iie	435	Ini	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	AAG Lys	1344
IIe.	450	GIY	Arg	Cys	Asp	455	Ala	Leu	Asn	Asp	Leu 460	Cys	Ile	Lys	GTT Val	1392
465	ASI	lip	Asp	ьeu	470	TTT Phe	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	Asn 480	1440
Asp	ren	ASD	гÄг	Gly 485	Glu	GAA Glu	Ile	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495	Ala	1488
Ala,	GIU	GIU	500	IIe	Ser	TTA Leu	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	Thr	1536
rne .	Asn	9ne 515	Asp	Asn	Glu	CCT Pro	Glu 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Leu	Ser	1584
ser .	530	ile	IIe	GIY	GIn	TTA Leu 535	Glu	Leu	Met	Pro .	Asn 540	Ile.	Glu	Arg	Phe	1632
CCT I Pro I 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT . Asp	Lys '	TAT / Tyr ' 555	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680

- 82 -

CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr		1728
AAT Asn	TCT Ser	GTT Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe		1776
TTT Phe	TCT Ser	TCA Ser 595	GAC Asp	TAT Tyr	GTA Val	AAG Lys	AAA Lys 600	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala		1824
ATG Met	TTT Phe 610	TTA Leu	GGC Gly	TGG Trp	GTA Val	GAA Glu 615	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GAA Glu		1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640		1920
ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys		1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu		2016
GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	•	2064
GTA Val	TCA Ser 690	TAT Tyr	ATT Ile	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA Ile	GAT Asp	AAT Asn		2112
GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720		2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg		2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala		2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn		2304
ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile		2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800		2400
										GGT Gly							2448
										TTA Leu							2496

AAT Asn	AGA Arg	GGA Gly 835	Thr	TTA Leu	ATT	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AAT Asn	AAT Asn 850	Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val		2592
GAT Asp 865	Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT	ACT Thr	GAA Glu 875	TAT	ATT Ile	AAG Lys	TCA Ser	GGC Gly 880		2640
CTG Leu	AAT Asn	TCC	CCG Pro	GGT Gly 885	Ala	GCT Ala	CAT	TAT	GCG Ala 890	CAA Gln	CAC His	GAT Asp	GAA Glu	GCC Ala 895	GTA Val		2688
GAC Asp	AAC Asn	AAA Lys	TTC Phe 900	AAC Asn	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT Tyr	GAG Glu 910	ATC Ile	TTA Leu		2736
CAT His	TTA Leu	CCT Pro 915	Asn	TTA Leu	AAC Asn	GAA Glu	GAA Glu 920	CAA Gln	CGA Arg	AAC Asn	GCC Ala	TTC Phe 925	ATC Ile	CAA Gln	AGT Ser		2784
TTA Leu	AAA Lys 930	GAT Asp	GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys		2832
AAG Lys 945	CTA Leu	AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960		2880
Glu	Gln	Gln	Asn	Ala 965	Phe	TAT Tyr	Glu	Ile	970	His	Leu	Pro	Asn	Leu 975	Asņ		2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC Phe	ATC Ile	CAA Gln 985	AGT Ser	TTA Leu	AAA Lys	GAT Asp	GAC Asp 990	CCA Pro	AGC Ser		2976
CAA Gln	AGC Ser	GCT Ala 995	AAC Asn	CTT Leu	TTA Leu	GCA Ala	GAA Glu 1000	Ala	AAA Lys	AAG Lys	CTA Leu	AAT Asn 1005	Asp	GCT Ala	CAG Gln	-	3024
	CCG Pro 1010	Lys			TAG						,						3042

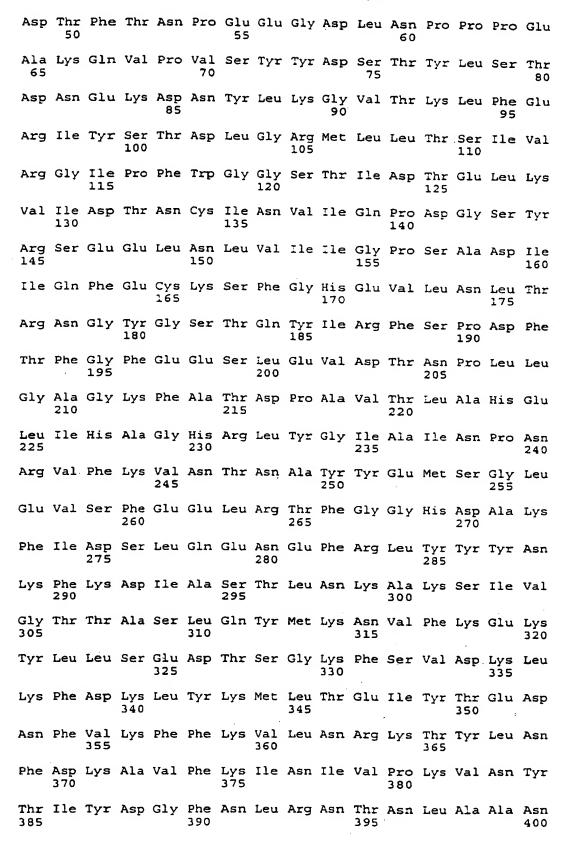
#### (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1014 amino acids
  - (B) TYPE: amino acid .
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 4 C



Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 630 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650 Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala

- 86 -

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 825 Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 905 His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 950 Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 Ala Pro Lys Val Asp \*

- (2) INFORMATION FOR SEQ ID NO: 19:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 3509 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:

1010

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..3509
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATO Met	Pro	A GTT	C ACA L Thr	ATA Ile	AAT Asr	TAA T Asn	TTT Phe	AAT Asn	TAT	Asr	GAI Asp	CC1	T ATT	GAT Asp	TAAT Asn	<i>:</i>	48
AAT Asn	TAAT AST	T ATT	T ATT e Ile 20	ATC Met	ATC Met	GAG Glu	CCT	CCA Pro 25	Phe	GCC Ala	AGA Arg	GG1 Gl <sub>3</sub>	ACG Thr	Gly	AGA Arg		96
TAT Tyr	TAI	AAA Lys 35	: Ala	TTT Phe	Lys	ATC Ile	ACA Thr 40	Asp	CGI Arg	T ATT	Trp	ATA Ile	Ile	CCC Pro	GAA Glu		144
AGA Arg	TAT Tyr 50	Thr	Phe	GGA Gly	TAT	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGT Ser	TCC	GGT		192
ATT Ile	Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT	GAT Asp 75	CCA Pro	GAT Asp	TAC	TTA Leu	AAT Asn 80	-	240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	Thr	ATG Met	ATC	AAG Lys	TTA Leu 95	Phe		288
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA	TTA Leu	GAG Glu 110	ATG Met	ATT Ile	-	336
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	Leu	GAA Glu	GAG Glu		384
TTT Phe	AAC Asn 130	ACA	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn		432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160		480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Ģlu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln		576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu		624
Asn	AAA Lys 210	GGC Gly	GCA : Ala	AGT Ser	Ile	TTT Phe 215	AAT . Asn .	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro		672
GCC Ala 225	TTG Leu	ATA Ile	TTA . Leu !	Met	CAT His 230	GAA ( Glu	CTT . Leu	ATA Ile	His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA. Leu	TAT Tyr 240		720
GGC Gly	ATT Ile	AAA Lys	Val A	GAT ( Asp 2	GAT Asp	TTA ( Leu	CCA /	Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe		768
TTT Phe	ATG Met	Gln	TCT A Ser 3 260	ACA (	GAT (	GCT A Ala :	Ile (	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe		816

- 88 -

Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 445  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  TTA TTA TTA GAT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT 500  TTA TTA TTA TCT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT 500  TTA TTA TTA TCT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT 500									-								
Tyr Asp Lys Val Leu Gl Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295   AAG GTT TTA GTT TGC ATA TCA GAT CCT AAC ATT AAT ATA TAT TAT Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 315   AAA AAT AAA TTT AAA GAT AAA TAT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 315   AAA AAT AAA TTT AAA GAT GAA AAT TT AAA TTC GTT GAA GAT TCT GAG GGA Lys Asn Lys Phe Lys Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340   ATG TTT GGT TTT ACA GAA ACT ATA TAT ACA GAA AAT TAT AAA ACC TTA Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345   AAG TTT GGT TTT ACA GAA ACT ATA TAT GCA GAA AAT TAT AAA ATA AAA AAAA	GGA Gly	GGA Gly	Gln	GAT Asp	CCC Pro	AGC Ser	ATC Ile	Ile	ACT Thr	CCT Pro	TCT Ser	ACG Thr	Asp	AAA Lys	AGT Ser	ATC Ile	864
195   195   196   197   197   198	TAT Tyr	Asp	AAA Lys	GTT Val	TTG Leu	CAA Gln	Asn	TTT Phe	AGA Arg	GGG Gly	ATA Ile	Val	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
Lys Asn Lys Phe Lye Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325  AAA TAT AGT ATA GAT GTA GAA AGT TTT GAT AAA TTA TAT AAA AGC TTA 340 AGT TYR Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340  ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA ATA AAA Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355  ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA ATA AAA THA ARA GAT ARA ATA ARA SER Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375  AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA AAA ATA AAA ATA AAA GAT TAT ARA AGA GAT TAT ACT ATA GAG GAT GAT TAT ATA AAA GAT ATA AAA AA	Lys	GTT Val	TTA Leu	GTT Val	TGC Cys	Ile	TCA Ser	GAT Asp	CCT Pro	AAC Asn	Ile	AAT Asn	ATT Ile	AAT Asn	ATA Ile	Tyr	960
Lys Tyr Ser 11e Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340  ATG TTT GGT TTT ACA GAA ACT AAT ATA GCA GAA AAT TAT AAA ATA AAA TE STE SER LYS Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355  ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA THA AGA SER ASP SER Leu Pro Pro Val Lys Ile Lys 370  ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA TTA AGA SER ASP SER Leu Pro Pro Val Lys Ile Lys 370  AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA ATA ASN Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385  CTCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GAG GGT CCA AAT AAA GCT ATA 390  CTCT GAT AAA GAT ATG GAA AAA GAA ATT AGA GAG GAG	AAA Lys	AAT Asn	AAA Lys	TTT Phe	Lys	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	Phe	GTT Val	GAA Glu	GAT Asp	TCT Ser	Glu	GGA Gly	1008
Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 365  ACT AGA GCT TCT TAT TTT AGT GAT TCC TTA CCA CCA GTA AAA ATA AAA 1152  Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370  AAT TTA TTA GAT AAT GAA ATC TAT ACT ATA GAG GAA GGG TTT AAT ATA ASN Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 400  TCT GAT AAA GAT ATG GAA AAA GAA TAT AGA GAG GGT CAG AAT AAA GCT ATA Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405  AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT ATA AAT AAA CAA GCT TAT GAA GAA ATT TY Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 429  AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 450  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA 450  GAT GAT TTA TCT AAA AAC GAA AGA ATG GAA ATA GAA AAT AGT TTT TCA 450  GAT GAT TTA TCT AAA AAC GAA AGA ATG GAA ATA AAT AGT TTT TCA 450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAC AGT AAT AGA AAT AGA ATA AGA AAT AGA ATA	AAA Lys	TAT Tyr	AGT Ser	Ile	GAT Asp	GTA Val	GAA Glu	AGT Ser	Phe	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	Lys	AGC Ser	TTA Leu	1056
The Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Pro Val Lys Ile Lys 375 375 375 375 375 375 375 375 380 Val Lys Ile Lys 375 375 375 375 375 375 380 Val Lys Ile Lys 326 375 375 375 375 375 375 375 375 375 375	ATG Met	TTT Phe	Gly	TTT Phe	ACA Thr	GAA Glu	ACT Thr	Asn	ATA Ile	GCA Ala	GAA Glu	AAT Asn	Tyr	AAA Lys	ATA Ile	AAA Lys	1104
Ash Leu Leu Asp Ash Glu Ile Tyr Thr Ile Glu Gly Phe Ash Ile 400  TCT GAT AAA GAT ATG GAA AAA GAA TAT AGG GGT CAG AAT AAA GCT ATA AGG AGF ASP Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Ash Lys Ala Ile 415  AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT AGA Lys Glu Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 425  AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATT TTG GCT GTA TAT AGG AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATT TG ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 445  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AGT TTT TCA ASP ASP Leu Ser Lys ASP Glu Arg Ile Glu Tyr Ash Thr Gln Ser Ash 465  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT ASP ASP Leu Ser Lys ASP Glu Arg Ile Glu Tyr Ash Thr Gln Ser Ash 480  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA AAT ACA CAG AGT AAT ASP ASP Leu Ser Lys ASP Phe Pro Ile Ash Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA GAA AAT GAA TTA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Soo ASP ASP Leu Ser Lys Ile Glu Leu Pro Ser Glu Ash Thr Glu Ser Leu Thr 500  GAT TTT AAT GAA AAA GAA GTT CCA GTA TAT GAA AAT ACA GAA TCA CTT ACT Soo ASP ASP ASP ASP ASP CASP TO SER GLU ASP THR GIU SER LEU Thr 500  GAT TTT AAT GAA AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Soo ASP ASP ASP ASP ASP THE GIU Leu Pro Ser Glu Ash Thr Glu Ser Leu Thr 500  GAT TTT AAT GAA AAA ASP CAA CAA CATC TTT CAA TAT TTA TAC CTC CAG L532  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG L532  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG L532  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG CLT L532  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG CLT L532  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG CLT CAG CLT CTT CA	ACT Thr	Arg	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	Ser	GAT Asp	TCC Ser	TTA Leu	CCA Pro	Pro	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 415  AAT AAA CAA GCT TAT GAA GAA ATT AGC AAG GAG CAT TTG GCT GTA TAT Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420  AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435  GTT GAT AAT GAA AGT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA 450  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA 450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT GAT AAT ACA CAG AGT AAT ASP Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GAT GAT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA ASP Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	Asn	TTA Leu	TTA Leu	GAT Asp	AAT Asn	Glu	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	Glu	GAA Glu	GGG Gly	TTT Phe	AAT Asn	Ile	1200
Ash Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr  420  AAG ATA CAA ATG TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT ATT GAT Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA Val Asp Asn Glu Asp Leu Phe Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ACT TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	TCT Ser	GAT Asp	AAA Lys	GAT Asp	Met	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	Arg	GGT Gly	CAG Gln	AAT Asn	AAA Lys	Ala	Ile	1248
Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp  GTT GAT AAT GAA GAT TTG TTC TTT ATA GCT GAT AAA AAT AGT TTT TCA  Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser  450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	AAT Asn	AAA Lys	CAA Gln	Ala	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	Ser	AAG Lys	GAG Glu	CAT His	TTG Leu	Ala	GTA Val	TAT Tyr	1296
Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450  GAT GAT TTA TCT AAA AAC GAA AGA ATA GAA TAT AAT ACA CAG AGT AAT Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	AAG Lys	ATA Ile	Gln	ATG Met	TGT Cys	AAA Lys	AGT Ser	Val	AAA Lys	GCT Ala	CCA Pro	GGA Gly	Ile	TGT Cys	ATT Ile	GAT Asp	1344
Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470  TAT ATA GAA AAT GAC TTC CCT ATA AAT GAA TTA ATT TTA GAT ACT GAT TYr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA ASP Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	GTT Val	Asp	AAT Asn	GAA Glu	GAT Asp	TTG Leu	Phe	TTT Phe	ATA Ile	GCT Ala	GAT Asp	Lys	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 495  TTA ATA AGT AAA ATA GAA TTA CCA AGT GAA AAT ACA GAA TCA CTT ACT Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA ASp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	Asp	GAT Asp	TTA Leu	TCT Ser	AAA Lys	Asn	GAA Glu	AGA Arg	ATA Ile	GAA Glu	Tyr	AAT Asn	ACA Thr	CAG Gln	AGT Ser	Asn	1440
Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500  GAT TTT AAT GTA GAT GTT CCA GTA TAT GAA AAA CAA CCC GCT ATA AAA Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	TAT Tyr	ATA Ile	GAA Glu	AAT Asn	Asp	TTC Phe	CCT Pro	ATA Ile	AAT Asn	Glu	TTA Leu	ATT Ile	TTA Leu	GAT Asp	Thr	GAT Asp	1488
Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525  AAA ATT TTT ACA GAT GAA AAT ACC ATC TTT CAA TAT TTA TAC TCT CAG Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	TTA Leu	ATA Ile	AGT Ser	Lys	ATA Ile	GAA Glu	TTA Leu	CCA Pro	Ser	GAA Glu	AAT Asn	ACA Thr	GAA Glu	Ser	CTT Leu	ACT Thr	1536
Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln	GAT Asp	TTT Phe	Asn	GTA Val	GAT Asp	GTT Val	CCA Pro	Val	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	Pro	GCT Ala	ATA Ile	AAA Lys	1584
	AAA Lys	Ile	TTT Phe	ACA Thr	GAT Asp	GAA Glu	Asn	ACC Thr	ATC Ile	TTT Phe	CAA Gln	Tyr	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

- 89 -

		_															
ACA Thi 545	Phe	CC.	r CTA	A GAT 1 Asp	11e 550	Arg	GAT Asp	ATA	AGT Ser	TTA Leu 555	Thi	TC1	TCA Ser	TTT Phe	GAT Asp 560		1680
GA1 Asp	GCA Ala	t TT/	A TTA	TTT Phe 565	Ser	AAC Asn	Lys	GTT Val	TAT Tyr 570	Ser	TTT Phe	TTT Phe	TCT Ser	ATC Met 575	GAT Asp		1728
TAT Tyr	ATT	AAA Lys	A ACT 5 Thr 580	Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	Glu	GCA Ala	GGA Gly	TTA Leu	TTI Phe 590	: Ala	GGT		1776
TGC Trp	GTG Val	Lýs 595	A CAG S Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	Phe	GTA Val	ATC	GAA Glu	GCT Ala 605	Asn	AAA Lys	AGC Ser		1824
AAT Asn	ACT Thr 610	Met	GAT Asp	AAA Lys	ATT	GCA Ala 615	GAT Asp	ATÀ Ile	TCT	CTA Leu	ATT Ile 620	Val	CCT Pro	TAT	ATA Ile		1872
GGA Gly 625	Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	Lys	GGA Gly	AAT Asn	TTT Phe	GAA Glu 640		1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro		1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA	GAA Glu	TCA Ser 670	TAT	ATT Ile	-	2016.
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys	-	2064
AGA Arg	AAT Asn 690	Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG Trp		2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720		2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr		2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp		2256
TTT Phe	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile		2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	TTT Phe	ATA Ile 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met		2352
AAA Lys 785	AAA Lys	ATG Met	ATT Ile	Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800		2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA . Leu .	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr		2448

reu	116	GGA Gly	820	Ala	GIU	Tyr	Glu	Lys 825	Ser	Lys	Val	Asn	Lys 830	Tyr	Leu	249	96
AAA Lys	ACC	Ile 835	Met	Pro	Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	254	14
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	Leu	AAT Asn	AAT Asn	ATT Ile	259	92
ATC Ile 865	TTA Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	AAG Lys	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880	264	0
TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GAG Glu	GTA Val	TAT	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys	268	8
AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr	273	6
CAA Gln	AAT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Lėu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val	278	4
AGC Ser	TTT Phe 930	TGG Trp	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn	283	2
TAT Tyr 945	ATT Ile	CAT His	AAT Asn	GAA Glu	TAT Tyr 950	ACA Thr	ATA Ile	ATT Ile	AAT Asn	TGT Cys 955	ATG Met	AAA Lys	AAT Asn	AAT Asn	TCG Ser 960	288	0
GGC Gly	TGG Trp	AAA Lys	ATA Ile	TCT Ser 965	ATT	AGG Arg	GGT Gly	AAT Asn	AGG Arg 970	ATA Ile	ATA Ile	TGG Trp	ACT Thr	TTA Leu 975	ATT Ile	292	8
GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	TCG Ser	GTA Val 985	TTT Phe	TTT Phe	GAA Glu	TAT Tyr	AAC Asn 990	ATA Ile	AGA Arg	297	6
GAA Glu	GAT Asp	ATA Ile 995	TCA Ser	GAG Glu	TAT Tyr	ATA Ile	AAT Asn 1000	Arg	TGG Trp	TTT Phe	TTT Phe	GTA Val 1005	Thr	ATT Ile	ACT Thr	3024	4
AAT Asn	AAT Asn 1010	TTG Leu	AAT Asn	AAC Asn	Ala	AAA Lys 1015	Ile	TAT Tyr	ATT Ile	AAT Asn	GGT Gly 1020	Lys	CTA Leu	GAA Glu	TCA Ser	307:	2
AAT Asn 1025	Thr	GAT Asp	ATT Ile	AAA Lys	GAT Asp 1030	Ile	AGA Arg	GAA Glu	GTT Val	ATT Ile 1035	Ala	TAA neA	GGT Gly	GAA Glu	ATA Ile 1040	3120	0
ATA Ile	TTT Phe	AAA Lys	TTA Leu	GAT Asp 1045	Gly	GAT Asp	ATA Ile	GAT Asp	AGA Arg 1050	Thr	CAA Gln	TTT Phe	ATT Ile	TGG Trp 1055	Met	3168	8
AAA Lys	TAT Tyr	TTC Phe	AGT Ser 1060	Ile	TTT Phe	AAT Asn	Thr	GAA Glu 1065	Leu	AGT Ser	CAA Gln	TCA Ser	AAT Asn 1070	Ile	GAA Glu	3216	6
GAA Glu	AGA Arg	TAT Tyr 1075	Lys	ATT Ile	CAA Gln	Ser	TAT Tyr 1080	Ser	GAA Glu	TAT Tyr	TTA Leu	AAA Lys 1085	Asp	TTT Phe	TGG Trp	3264	4

- 91 -

GGA Gly	AAT CO Asn Pi 1090	CT TTA ro Leu	ATG TA Met Ty	C AAT r Asn 1095	Lys Gl	A TAT lu Tyr	TAT Tyr	ATG Met 1100	Phe	AAT Asn	GCG Ala	GGG Gly	3312	2
AAT Asn 110	Lys As	AT TCA sn Ser	TAT AT Tyr Il	e Lys	CTA AA Leu Ly	AG AAA 7s Lys.	GAT Asp 1115	Ser	CCT Pro	GTA Val	GGT Gly	GAA Glu 1120	3360	)
ATT Ile	TTA AC	CA CGT hr Arg	AGC AA Ser Ly: 1125	A TAT	AAT CA Asn Gl	A AAT n Asn 1130	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1139	Tyr	3408	3
AGA Arg	GAT TT Asp Le	TA TAT Eu Tyr 1140	ATT GG     Ile Gl	A GAA /	Lys Ph	T ATT e Ile 45	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	3456	;
TCT Ser	Gln Se	T ATA er Ile	AAT GAT Asn Ast	Asp :	ATA GT Ile Va 1160	T AGA 1 Arg	AAA Lys	Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3504	ř
CTA Leu	.GA								•				3509	ı

#### (2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1169 amino acids
  - (B) TYPE: amino acid
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe

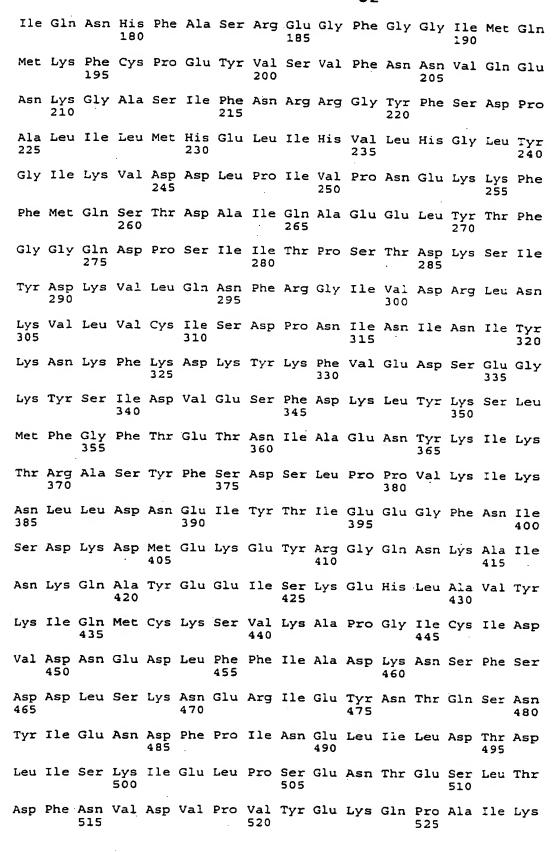
Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 135

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165



Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Giu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 650 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 730 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 790 Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 810 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 840 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly

- 94 -

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 920 Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 1000 Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1015 1010 Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1050 Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1065 Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1095 Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1110 1115 Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1145 Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1160

Leu

- (2) INFORMATION FOR SEQ ID NO: 21:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)

- 95 -

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Pro	GTI Val	ACA Thr	ATA Ile	AAT Asn	AAT Asn	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT	ATT Ile	GAT Asp 15	AAT Asn		48
'AAT Asn	ATT	Ile	Met	ATG Met	GAG Glu	CCT Pro	Pro	Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG	AGA Arg		96
TAT	Lys	Ala	TTT	AAA Lys	ATC	ACA Thr 40	GAT	CGT	ATT Ile	TGG Trp	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu		144
TAT Tyr 50	Thr	TTT Phe	GGA Gly	TAT	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	AAA Lys	AGT Ser	TCC Ser	GGT Gly		192
Phe	AAT Asn	. AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	TAA nsA 08		240
AAT Asn	GAT	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe	•	288
AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT. Ile		336
AAT Asn	GGT Gly 115	ATA Ile	CCT	TAT	CTT	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu		384
AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn		432
GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160		480
GGA Gly	CCT Pro	GGG	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
CAA Gln	AAT Asn	CAT His 180	TTT	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln		576
AAG Lys	TTT Phe 195	TGC ·	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu		624
AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro		672
TTG Leu	ATA Ile	TTA Leu	ATG Met	His	Glu	CTT Leu	ATA Ile	His	Val	TTA Leu	CAT His	GGA Gly	Leu	Tyr		720
	AAT AS AT TYP SO TITT Phe AAT AS A AS A AS A GIY GGA GIY CAA AGIN AAG LYS AAA LYS 210 TTG	AAT ATT AAA TYT Lys 35 TAT ACT TYT Thr 50 TTT AAT ASN	AAT ATT ATT ATT AS I I I I I I I I I I I I I I I I I I	AAT ATT ATT ATG ASN Ile Ile Met 20  TAT AAA GCT TTT Tyr Lys Ala Phe 35  TAT ACT TTT GGA Tyr Thr Phe Gly 50  AAT GAT AAA AAG ASN ASP Lys Lys 85  AGA ATC AAA TCA Arg Ile Lys Ser 100  AAT GGT ATA CCT ASN Gly Ile Pro 115  AAC ACA AAC ATT ASN Thr ASN Ile 130  GGA GAA GTG GAG Gly Glu Val Glu  GGA CCT GGG CCA Gly Pro Gly Pro 165  CAA AAT CAT TTT Gln ASN His Phe 180  AAG TTT TGC CCA Lys Phe Cys Pro 195  AAA GGC GCA AGT Lys Gly Ala Ser 210  TTG ATA TTA ATG	AAT ATT ATT ATG ATG ASN Ile Ile Met Met 20  TAT AAA GCT TTT AAA TYR Lys Ala Phe Lys 35  TAT ACT TTT GGA TAT TYR Thr Phe Gly TYR 50  TTT AAT AGA GAT GTT Phe ASN ARG AAT ASN ASP Lys Lys ASN 85  AGA ATC AAA TCA AAA ARG AAT ASN ASP Lys Lys Ser Lys 100  AAT GGT ATA CCT TAT ASN Gly Ile Pro Tyr 115  AAC ACA AAC ATT GCT ASN Thr ASN Ile Ala 130  GGA GAA GTG GAG CGA GIY Glu Arg 150  GGA CCT GGG CCA GTT Gly Pro Gly Pro Val 165  CAA AAT CAT TTT GCA GAA Lys Phe Cys Pro Glu 195  AAA GGC GCA AGT ATA LYS GRA GIY GIY AIS Phe Cys Pro Glu 195  AAA GGC GCA AGT ATA TTT TTT GCA TATA TTT THE AIS TTTT THE AIS TATA ATG CAT LEU IIE LEU MET HIS	Pro Val Thr Ile Asn Asn S  AAT ATT ATT ATG ATG GAG Asn Ile Ile Met Met Glu 20  TAT AAA GCT TTT AAA ATC Tyr Lys Ala Phe Lys Ile 35  TAT ACT TTT GGA TAT AAA Tyr Thr Phe Gly Tyr Lys 50  TTT AAT AGA GAT GTT TGT Phe Asn Arg Asp Val Cys 70  AAT GAT AAA AAG AAT ATA ASn Asp Lys Lys Asn Ile 85  AGA ATC AAA TCA AAA CCA Arg Ile Lys Ser Lys Pro 100  AAT GGT ATA CCT TAT CTT Asn Gly Ile Pro Tyr Leu 115  AAC ACA AAC ATT GCT AGT Asn Thr Asn Ile Ala Ser 130  GGA GAA GTG GAG CGA AAA Gly Glu Val Glu Arg Lys 150  GGA CCT GGG CCA GTT TTA GLY GIY Pro Gly Pro Val Leu 165  CAA AAT CAT TTT GCA TCA Gln Asn His Phe Ala Ser 180  AAG TTT TGC CCA GAA TAT Lys Phe Cys Pro Glu Tyr 195  AAA GGC GCA AGT ATA TTT Lys Gly Ala Ser Ile Phe 210  TTG ATA TTA ATG CAT GAA	AAT ATT ATT ATG ATG GAG CCT ASN Ile Ile Met Met Glu Pro 20  TAT AAA GCT TTT AAA ATC ACA TYR Lys Ala Phe Lys Ile Thr 35  TAT AAT ACT TTT GGA TAT AAA CCT TYR Thr Phe Gly TYR Lys Pro 50  TAT AAT AGA GAT GTT TGT GAA Phe ASN ARG ASP Val Cys Glu 70  AAT GAT AAA AAG AAT ATA TTT ASN ASP Lys Lys Pro Leu 100  AAT GGT ATA CCT TAT CTT GGA ASN Gly Ile Pro Tyr Leu Gly 115  AAC ACA AAC ATC TYR GAT GTA AAA CAR GRA THR ASN Ile Ala Ser Val 130  GGA GAA GTG GAG CGA AAA AAA Gly Glu Val Glu Arg Lys Lys Lys GGA CCT GGG CCA GTT TTA AAT Gly Pro Gly Pro Val Leu Asn 165  CAA AAT CAT TTT GCA TCA AGG Gln ASN His Phe Ala Ser Arg 180  AAG GTT TGC CCA GAA TAT GTA Lys Phe Cys Pro Glu Tyr Val 195  TTG ATA TTA ATG CAT GAA CTT CTT GAA CTT Leu Ile Leu Met His Glu Leu	AAT ATT ATT ATG ATG GAG CCT CCA ASN Ile Ile Met Met Glu Pro Pro 20 25 TAT AAA GCT TTT AAA ATC ACA GAT Tyr Lys Ala Phe Lys Ile Thr Asp 35 AAT ATT ATT GGA TAT AAA CCT GAG Tyr Thr Phe Gly Tyr Lys Pro Glu 50 TTT AAT AGA GAT GTT TGT GAA TAT Phe Asn Arg Asp Val Cys Glu Tyr 70 AAT GAT AAA AAG AAT ATA TTT TTA ASN ASP Lys Lys Asn Ile Phe Leu 85 AGA ATC AAA TCA AAA CCA TTG GGT Arg Ile Lys Ser Lys Pro Leu Gly 100 AAT GGT ATA CCT TAT CTT GGA GAT Asn Gly Ile Pro Tyr Leu Gly Asp 115 AAC ACA AAC ATT GCT AGT GTA ACT ASN Thr Asn Ile Ala Ser Val Thr 130 GGA GAA GTG GAG CGA AAA AAA GGT Gly Glu Val Glu Arg Lys Lys Gly 150 GGA CCT GGG CCA GTT TTA AAT GAA GIV Pro Gly Pro Val Leu Asn Glu 165 CAA AAT CAT TTT GCA TCA AGG GAA GIN ASN His Phe Ala Ser Arg Glu 180 AAG GTT TGC CCA GAA TAT GTA AGC Lys Phe Cys Pro Glu Tyr Val Ser 195 AAAA GGC GCA AGT ATA TTT AAT AGA Lys Gly Ala Ser Ile Phe Asn Arg 210 TTG ATA TTA ATG CAT GAA CTT ATA Leu Ile Leu Met His Glu Leu Ile	Pro Val Thr Ile Asn Asn Phe Asn Tyr 10  AAT ATT ATT ATG ATG GAG CCT CCA TTT Asn Ile Ile Met Met Glu Pro Pro Phe 20  TAT AAA GCT TTT AAA ATC ACA GAT CGT Tyr Lys Ala Phe Lys Ile Thr Asp Arg 40  TAT ACT TTT GGA TAT AAA CCT GAG GAT Tyr Thr Phe Gly Tyr Lys Pro Glu Asp 55  TTT AAT AGA GAT GTT TGT GAA TAT TAT Phe Asn Arg Asp Val Cys Glu Tyr Tyr 70  AAT GAT AAA AAG AAT ATA TTT TTA CAA Asn Asp Lys Lys Asn Ile Phe Leu Gln 90  AGA ATC AAA TCA AAA CCA TTG GGT GAA Arg Ile Lys Ser Lys Pro Leu Gly Glu 105  AAT GGT ATA CCT TAT CTT GGA GAT AGA Ang Ile Pro Tyr Leu Gly Asp Arg 115  AAC ACA AAC ATT GCT AGT GTA ACT GTT Asn Thr Asn Ile Ala Ser Val Thr Val 130  GGA GAA GTG GAG CGA AAA AAA GGT ATT GIY GAA AAT GGT ATT GIY Asn Thr Asn Ile Ala Ser Val Thr Val 150  GGA CCT GGG CCA GTT TTA AAT GAA AAT GGT ATT GIY GIY Asn Ile Ala Ser Yal Thr Val 165  CAA AAT CAT TTT GCA TCA AGG GAA GGC GIY ASn His Phe Ala Ser Arg Glu Gly 185  AAG TTT TGC CCA GAA TAT GTA AGG GAA GGC GTA AGA GTT TGC CCA GAA TAT TTT AAT AGC GTA AGA GIY AIS Phe Cys Pro Glu Tyr Val Ser Val 195  AAA GGC GCA AGT ATA TTT AAT AGA CGT ATT Lys Phe Cys Pro Glu Tyr Val Ser Val 195  AAA GGC GCA AGT ATA TTT AAT AGA CGT ATT Lys Gly Ala Ser Ile Phe Asn Arg Arg Arg Arg Clu Gly Ala Ser Ile Phe Asn Arg Arg Arg Arg Clu Ile Leu Met His Glu Leu Ile His	AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG ASN Ile He Met Met Glu Pro Pro Phe Ala 20 25  TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TYR Lys Ala Phe Lys Ile Thr Asp Arg Ile 35  TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT TYR Thr Phe Gly Tyr Lys Pro Glu Asp Phe 55  TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp 70 75  AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ASN Asp Lys Lys Asn Ile Phe Leu Gln Thr 85  AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG ARG Ile Lys Ser Lys Pro Leu Gly Glu Lys 105  AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT ASN Gly Ile Pro Tyr Leu Gly Asp Arg Arg 115  AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT ASN Thr ASN Ile Ala Ser Val Thr Val Asn 130  GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GIY Gly Glu Val Glu Arg Lys Lys Gly Ile Phe 150  GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG GTT CTA ASN His Phe Ala Ser Arg Glu Asn Glu 165  CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GIN ASN His Phe Ala Ser Arg Glu Gly Phe 180  AAG TTT TGC CCA GAA TAT GTA AGG GAA GGC TTC GIN ASN His Phe Ala Ser Arg Glu Gly Phe 180  AAA GGC GCA AGT ATA TTT TAT GAA AGA GGT ATT TTC CLA AAT CAT TTT TTT TAT CAA AAT GAG GTA TTT TTC GAA AAT CAT TTT TTT TAT CAT GTT AAT ASN His Phe Ala Ser Arg Glu Gly Phe 180  AAA GGC GCA AGT ATA TTT AAT GAA GGC GTA TTT TTC CLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT CAT TTT GCA TCA AGG GAA GGC TTC GLA AAT GAA GAT ATT TTC GAA AAT TTT AAT AGA CAT GTT TTC GAA AAT TTA AAT AGA CAT GAT GTT AAT AAT AAT AAT AAT AAT GAA AAT GTT GT	AAT ATT ATT GGA TAT AAA ATC ACA GAT CGT ATT ASP	Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro S 10   AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ASN Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly 25   TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA TYR Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Asn ACT ACT TTT GGA TAT AAA ACT ACT GAG GAT TTT AAT AAA TYR Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys 50   TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp 70   AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC ASN ASp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile 85   AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA ARG GIV Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro 120   AAT GAT AAA AS ATT GCT TAT CTT GGA GAT AGA CGT GTT CCA Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro 120   AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ASN THA ASN Ile Ala Ser Val Thr Val Asn Lys Leu 130   AGGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT GLY Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn 155   AGA ACT GGG CCA GTT TTA AAT GAA ACT GAT TTC GGA GAT ATT TTC GCA AAT GLY Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn 150   AAC ACA AAC ATT GCT TAT CTT GGA GAT ATT TTC GCA AAT GLY Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn 150   AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ASN THA ASN Ile Ala Ser Arg Gly Ile Phe Ala Asn 150   AAC CT GGG CCA GTT TTA AAT GAA ACT GTT TTC GCA AAT GLY Fro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile 165   AAA TTT TGC CCA GAA TAT GTA AGG GAA GGC TTC GGG GGT Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly 185   AAA GGC GCA AGT ATA TTT AAT AGG CGT GTA TTT AAT AAT Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn 195   AAA GGC GCA AGT ATA TTT AAT AGG CGT GGA TAT TTT Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe 210   AAA GGC GCA AGT ATA TTT AAT AGA CTT TTA CAT Leu His Val Leu His	Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile  AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr 20  TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile 45  TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser 50  TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr 70  AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys 85  AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TTA GAG Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu 100  AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu 115  AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile 130  GGA GAA GTG GAG CGA AAA AAA AAA GGT ATT TTC GCA AAT TTA GIY Glu Val Glu Arg Lys Lys Gly Ile Phe Asn Leu 165  GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ACT ATA GAT Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp 165  CAA AAT CAT TTT GCA TCA AGG GAA GGC TTC GGG GGT ATA Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile 180  AAA GGC CCA GGA ATA TTT AAT AAG GGT TTT AAT AAT AA	Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp 15  AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGT ASN Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly 30  TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro 15  TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser 50  TTT AAT AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu 75  AAT GAT AAA AAG AAT ATA TTT TTA CAA ACA ATG ATC AAG TTA Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu 90  AGA ATC AAA TCA AAA CCA TTG GGT GAA AAG TTA TATA GAT CAC CTC GAA Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Glu Met 105  AAT GGT ATA CCT TAT CTT GGA GAT AGA CGT GTT CCA CTC GAA ASn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu 115  AAC ACA AAC ATT GCT AGT GTA ACT GTT AAT AAA TTA ATC AGT Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser 130  GGA GAA GTG GAG CGA AAA AAA GGT ATT TTC GCA AAT TTA ATC AGT Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser 130  GGA CCT GGG CCA GTT TTA AAT GAA AAT GAG ATT ATA GAT ATA ATA	AAT ATT ATT ATG ATG GAG CCT CCA TTT GCG AGA GGT ACG GGA GAGA TY I Le Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 25    TAT AAA GCT TTT AAA ATC ACA GAT CGT ATT TGG ATA ATA CCG GAA TY Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 45    TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT TY THR Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50    TAT ACT TTT GGA TAT AAA CCT GAG GAT TTT AAT AAA AGT TCC GGT TY THR Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50    TAT AAA AGA GAT GTT TGT GAA TAT TAT GAT CCA GAT TAC TTA AAT Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn ATG	Pro Val Thr 11e Asn Asn Phe Asn Tyr Asn Asp Pro 11e Asp Asn 15 10 15 15 15 16 Asn Att Att Att Att Att Att Att Att Att At

GGC Gly	: ATT	AAA Lys	A GTA S Val	GAT Asp 245	Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT Phe	Met	CAA Gln	TCI Ser 260	inr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	816
GGA Gly	GGA Gly	CAA Gln 275	Asp	Pro	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	ьys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	vaı	TTA Leu	GTT Val	TGC	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

														. •			
GAT Asp	TTT	AAT Asn 515	Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT	ATA Ile	AAA Lys		1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC	TCT Ser	CAG Gln		1632
		CCT Pro															1680
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTT Phe 565	TCT Ser	AAC Asn	AAA Lys	GTT Val	TAT Tyr 570	TCA Ser	TTT Phe	TTT Phe	TCT Ser	ATG Met 575	GAT Asp		1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	GCT Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	GAA Glu	GCA Ala	GGA Gly	TTA Leu	TTT Phe 590	GCA Ala	GGT Gly		1776
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	TTT Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	AAT Asn	AAA Lys	AGC Ser		1824
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT Ile	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	GTT Val	CCT Pro	TAT Tyr	ATA Ile		1872
		GCT Ala															1920
		TTT Phe															1968
		TTA Leu		Pro													2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys		2064
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	Met	Tyr	GGA Gly	Leu	Ile	Val	GCG Ala	CAA Gln	TGG Trp	:	2112
		ACA Thr															2160
		TTA Leu			Gln											•	2208
		AAT Asn															2256
		GAT Asp 755															2304
Asp		ATA Ile															2352

- 98 -

AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
		GAA Glu									~					2574

# (2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 858 amino acids(B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

			-					-							
Met 1	Pro	Val	Thr	Ile 5	Asn	Asn	Phe	Asn	Tyr 10	Asn	Asp	Pro	Ile	Asp 15	Asn
Asn	Asn	Ile	Ile 20	Met	Met	Glu	Pro	Pro 25	Phe	Ala	Arg	Gly	Thr 30	Gly	Arg
Tyr	Tyr	Lys 35	Ala	Phe	Lys	Ile	Thr 40	Asp	Arg	Ile	Trp	Ile 45	Ile	Pro	Glu
Arg	Tyr 50	Thr	Phe	Gly	Tyr	Lys 55	Pro	Glu	Asp	Phe	Asn 60	Lys	Ser	Ser	Gly
Ile 65	Phe	Asn	Arg	Asp	Val 70	Cys	Glu	Tyr	Tyr	Asp 75	Pro	Asp	Tyr	Leu	Asn 80
Thr	Asn	qaA	Lys	Lys 85	Asn	Ile	Phe	Leu	Gln 90	Thr	Met	Ile	Lys	Leu 95	Phe
Asn	Arg	Ile	Lys 100	Ser	Lys	Pro	Leu	Gly 105	Glu	Lys	Leu	Leu	Glu 110	Met	Ile
Ile	Asn	Gly 115	Ile	Pro	Tyr	Leu	Gly 120	Asp	Arg	Arg	Val	Pro 125	Leu	Glu	Glu
Phe	Asn 130	Thr	Asn	Ile	Ala	Ser 135	Val	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn
Pro 145	Gly	Glu	Val	Glu	Arg 150	Lys	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160
Phe	Gly	Pro	Gly	Pro 165	Val	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile	Asp	Ile 175	Gly
Ile	Gln	Asn	His 180	Phe	Ala	Ser	Arg	Glu 185	Gly	Phe	Gly	Gly	Ile 190	Met	Gln

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 395 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 520 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 540

- 100 -

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 665 Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 840 Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser

### (2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1644 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

	1.44	, 54	Q02.							· ·	<b>J</b> .						
ATG Met 1	Pro	GTT Val	ACA Thr	ATA Ile 5	Asn	AAT Asn	TTT	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT	GAT Asp 15	AAT Asn		48
AAT Asn	AAT Asn	ATT Ile	ATT Ile 20	ATG Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG Gly	AGA Arg		96
TAT Tyr	TAT Tyr	AAA Lys 35	Ala	TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	GAT Asp	CGT	ATT Ile	TGG Trp	ATA Ile	Ile	CCG Pro	GAA Glu		144
AGA Arg	TAT Tyr 50	ACT Thr	TTT Phe	GGA Gly	TAT	AAA Lys 55	CCT	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	AAA Lys	AGT Ser	TCC	GGT Gly		192
				GAT Asp													240
				AAG Lys 85													288
				TCA Ser													336
				CCT												•	3 8.4:
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn		432
				GAG Glu													480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly		528
ATA Ile	CAA Gln	Asn	His	TTT Phe	Ala	Ser	Arg	Glu	Gly	Phe	Gly	Gly	Ile	Met	CAA Gln		<b>5</b> 76
				CCA Pro													624
				AGT Ser													672 -

- 102 -

GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AA'A Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TG <b>T</b> Cys	ATT Ile	GAT Asp	1344
		AAT Asn														1392
		TTA Leu														1440
		GAA Glu														1488

- 103 -

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632
		CCT Pro		_												1644

- (2) INFORMATION FOR SEQ ID NO: 24:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 548 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein

195

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 105 Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 150 155 Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu

200

205

Asn	Lys 210	Gly	Ala	Ser	Ile	Phe 215	Asn	Arg	Arg	Gly	Tyr 220	Phe	Ser	Asp	Pro
Ala 225	Leu	Ile	Leu	Met	His 230	Glu	Leu	Ile	His	Val 235	Leu	His	Gly	Leu	Tyr 240
Gly	Ile	Lys	Val	Asp 245	Asp	Leu	Pro	Ile	Val 250	Pro	Asn	Glu	Lys	Lys 255	Phe
Phe	Met	Gln	Ser 260	Thr	Asp	Ala	Ile	Gln 265	Ala	Glu	Glu	Leu	Tyr 270	Thr	Phe
Gly	Gly	Gln 275	Asp	Pro	Ser	Ile	Ile 280	Thr	Pro	Ser	Thr	Asp 285	Lys	Ser	Ile
Tyr	Asp 290	Lys	Val	Leu	Gln	Asn 295	Phe	Arg	Gly	Ile	Val 300	Asp	Arg	Leu	Asn
Lys 305	Val	Leu	Val	Cys	Ile 310	Ser	Asp	Pro	Asn	Ile 315	Asn	Ile	Asn	Ile	Tyr 320
Lys	Asn	Lys	Phe	Lys 325	Asp	Lys	Tyr	Lys	Phe 330	Val	Glu	Asp	Ser	Glu 335	Gly
•	-		340					345					350	Ser	
		355					360					365		Ile	
	370					375					380			Ile	
385			_		390					395				Asn	400
		_		405					410					Ala 415	
	-		420					425					430	Val	
-		435	•				440					445		Ile	
Val	Asp 450	Asn	Glu	Asp	Leu	Phe 455	Phe	Ile	Ala	Asp	Lys 460	Asn	Ser	Phe	Ser
Asp 465	_	Leu	Ser	Lys	Asn 470	Glu	Arg	Ile	Glu	Tyr 475	Asn	Thr	Gln	Ser	Asn 480
Tyr	Ile	Glu	Asn	Asp 485		Pro	Ile	Asn	Glu 490	Leu	Ile	Leu	Asp	Thr 495	Asp
Leu	Ile	Ser	Lys 500		Glu	Leu	Pro	Ser 505		Asn	Thr	Glu	Ser 510	Leu	Thr
_		515					520					525		Ile	
Lys	1le 530		Thr	Asp	Glu	Asn 535		Ile	Phe	Gln	Tyr 540	Leu	Tyr	Ser	Gln
Thr 545		Pro	Leu	l											

- 105 -

# (2) INFORMATION FOR SEQ ID NO: 25:

# (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2616 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

# (ix) FEATURE:

- (A) NAME/KEY: CDS
  (B) LOCATION:1..2616

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

		GTG Val								48
		GCC Ala 20	Tyr							96
		TTC Phe								144
		ACG Thr								192
		GTG Val								240
		AAG Lys							Glu	288
		TCC Ser 100								336
		CCA Pro							AAG Lys	384
		ACT Thr								432
		GAA Glu								480
		GAG Glu								528
		TAC Tyr 180								576

- 106 -

ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arģ	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT. Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

- 107 -

AGT Ser 465	Pro	TC/ Sei	A GAZ	A GAI 1 Asp	AAT AST 470	) Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	•	1440
ATT Ile	ACA Thr	TCI Ser	GAT Asp	Thr 485	Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495			1488
GAT Asp	TTA Leu	ATA Ile	CAA Glr 500	A CAA n Gln	TAT	TAT	TTA Leu	ACC Thr 505	Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	GAA Glu	CCT Pro		1536
Glu	Asn	11e 515	Ser	ATA Ile	Glu	Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu		1584
GAA Glu	Leu 530	Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT	GAG Glu		1632
TTA Leu 545	Asp	AAA Lys	TAT	ACT Thr	ATG Met 550	TTC Phe	CAT	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT	GAA Glu 560		1680
His	Gly	Lys	Ser	AGG Arg 565	Ile	Ala	Leu	Thr	Asn 570	Ser	Val	Asn	Glu	Ala 575	Leu		1728
Leu	Asn	Pro	Ser 580		Val	Tyr	Thr	Phe 585	Phe	Ser	Ser	Asp	Tyr 590	Val	Lys	•	1776
Lys	Val	Asn 595	Lys	GCT Ala	Thr	Glu	Ala 600	Ala	Met	Phe	Leu	Gly 605	Trp	Val	Glu		1824
Gln'	Leu 610	Val	Tyr	GAT Asp	Phe	Thr 615	Asp	Glu	Thr	Ser	Glu 620	Val	Ser	Thr	Thr		1872
Asp 625	Lys	Ile	Ala	GAT Asp	Ile 630	Thr	Ile	Ile	Ile	Pro 635	Tyr	Ile	Gly	Pro	Ala: 640		1920
Leu	Asn	Ile	Gly	AAT Asn 645	Met	Leu	Tyr	Lys	Asp 650	Asp	Phe	Val	Gly	Ala 655	Leu		1968
Ile	Phe	Ser	Gly 660	GCT Ala	Val	Ile	Leu	Leu 665	Glu	Phe	Įle	Pro	Glu 670	Ile	Ala		2016
Ile	Pro	Val 675	Leu	GGT Gly	Thr	Phe	Ala 680	Leu	Val	Ser	Tyr	Ile 685	Ala	Asn	Lys		2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA . Glu		2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	Ala	AAG Lys 720		2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile. 725	GAT Asp	CTA Leu	ATA . Ile <sub>.</sub>	Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	Leu		2208 .

#### - 108 -

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn		2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT	GAT Asp	GAT Asp		2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile		2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800		2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys		2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	-	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp		2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser		2592
					ATT Ile 870		TAA									ï	2616

#### (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids

  - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 Arg Val Phe Lys Val Asm Thr Asm Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 315 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Vai Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 455

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 505 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 615 Asp Lys Ile Ala Asp Ile Thr Ile Ile Iie Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 795 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 81C

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 855

Thr Phe Thr Glu Tyr Ile Lys

- (2) INFORMATION FOR SEQ ID NO: 27:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA	CCATCAACAA	CTTCAACTAC	AACGACCCGA	TCGACAACAA	CAACATCATC	60
ATGATGGAAC	CGCCGTTCGC	ACGTGGTACC	GGTCGTTACT	ACAAGGCTTT	CAAGATCACC	120
GACCGTATCT	GGATCATCCC	GGAACGTTAC	ACCTTCGGTT	ACAAACCTGA	GGACTTCAAC	180
AAGAGTAGCG	GGATTTTCAA	TCGTGACGTC	TGCGAGTACT	ATGATCCAGA	TTATCTGAAT	240
ACCAACGATA	AGAAGAACAT	ATTCCTTCAG	ACTATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACC	TTATCTTGGA	360
GATAGACGTG	TTCCACTCGA	AGAGTTTAAC	ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGÀAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTTATAC	ATGTTTTACA	TGGATTATAT	720
GGCATTAAAG	TAGATGATTT	ACCAATTGTA	CCAAATGAAA	AAAAATTTTT	TATGCAATCT	780
ACAGATGCTA	TACAGGCAGA	AGAACTATAT	ACATTTGGAG	GACAAGATCC	CAGCATCATA	840
ACTCCTTCTA	CGGATAAAAG	TATCTATGAT	AAAGTTTTGC	AAAATTTTAG	AGGGATAGTT	900
GATAGACTTA	ACAAGGTTTT	AGTTTGCATA	TCAGATCCTA	ACATTAATAT	TAATATATAT	960
TAAAATAAAA	TTAAAGATAA	ATATAAATTC	GTTGAAGATT	CTGAGGGAAA	ATATAGTATA	1020
GATGTAGAAA	GTTTTGATAA	ATTATAAA	AGCTTAATGT	TTGGTTTTAC	AGAAACTAAT	1080
ATAGCAGAAA	ATTATAAAAT	AAAAACTAGA	GCTTCTTATT	TTAGTGATTC	CTTACCACCA	1140
GTAÄAAATAA	AAAATTTATT	AGATAATGAA	ATCTATACTA	TAGAGGAAGG	GTTTAATATA	1200

TCTGATAAAG ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATGAAGAAA TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCTCCAG GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGTTTTT CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAGAAA ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ÀTACTGATTT	AATAAGTAAA	1500
ATAGAATTAC CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAAAAAC AACCCGCTAT	AAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACTCTC AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТАААААТАА	AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
AATGAGGGTA TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA TGAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
TCAATATATA CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

#### (2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:

  (A) LENGTH: 2574 base pairs

  (B) TYPE: nucleic acid

  (C) STRANDEDNESS: double

  (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

60	TAATATTATT	TTGATAATAA	AATGATCCTA	TTTTAATTAT	CAATAAATAA	ATGCCAGTTA
120	TAAAATCACA	ATAAAGCTTT	GGGAGATATT	GAGAGGTACG	CTCCATTTGC	ATGATGGAGC
180	GGATTTTAAT	ATAAACCTGA	ACTTTTGGAT	GGAAAGATAT	GGATAATACC	GATCGTATTT
240	ТТАСТТАААТ	ATCATCCACA	TCTCAATATT	TAGAGATGTT	ב מדיייייייייייי מ	

ACTAATGATA	AAAAGAATAT	ATTTTTACAA	ACAATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACC	TTATCTTGGA	360
GATAGACGT	TTCCACTCGA	AGAGTTTAAC	: ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTI	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTI	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATO	CAGCCTTGAT	ATTAATGCAT	GAACTCATCC	ACGTCCTCCA	CGGTCTCTAC	720
GGTATCAAAG	TAGACGACCT	CCCGATCGTC	CCGAACGAAA	AAAAATTCTT	CATGCAGAGC	780
ACCGACGCAA	TCCAGGCAGA	AGAACTCTAC	ACCTTCGGTG	GTCAGGACCC	GAGCATCATC	840
ACCCCGAGCA	CCGACAAAAG	CATCTACGAC	AAAGTCCTCC	AGAACTTCCG	TGGTATCGTC	900
GACCGTCTCA	ACAAAGTCCT	CGTCTGCATC	AGCGACCCGA	ACATCAACAT	CAACATCTAC	960
AAAAACAAAT	TCAAAGACAA	ATACAAATTC	GTCGAAGACA	GCGAAGGTAA	ATACAGCATC	1020
GACGTCGAGA	GCTTCGACAA	ACTCTACAAA	AGCCTCATGT	TCGGTTTCAC	CGAAACCAAC	1080
ATCGCAGAAA	ACTACAAAAT	CAAAACCCGT	GCAAGCTACT	TCAGCGACAG	CCTCCCGCCG	1140
GTCAAAATCA	AAAACCTCCT	CGACAACGAA	ATCTACACCA	TCGAAGAAGG	TTTCAACATC	1200
AGCGACAAAG	ACATGGAAAA	AGAATACCGT	GGTCAGAACA	AAGCAATCAA	CAAACAAGCT	1260
TACGAAGAAA	TCAGCAAAGA	ACACCTCGCA	GTCTACAAAA	TCCAGATGTG	CAAAAGCGTC	1320
AAAGCACCGG	GTATCTGCAT	CGACGTTGAC	AACGAAGACC	TCTTCTTCAT	CGCAGACAAA	1380.
AACAGCTTCA	GCGACGACCT	CAGCAAAAAC	GAACGTATCG	AATACAACAC	CCAGAGCAAC	1440
TACATCGAAA	ACGACTTCCC	GATCAACGAA	CTCATCCTCG	ACACCGACCT	CATCAGCAAA	1500
ATCGAACTCC	CGAGCGAAAA	CACCGAAAGC	CTCACCGACT	TCAACGTTGA	CGTCCCGGTC	1560
TACGAAAAAC	AGCCGGCAAT	CAAAAAAATC	TTCACCGACG	AAAACACCAT	CTTCCAGTAC	1620
CTCTACAGCC	AGACCTTCCC	GCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATAA	AATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	АТАТААТАТА	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280

PCT/GB97/02273

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
TCAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

#### **CLAIMS**

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain  $H_N$  portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain  $\beta$  of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated  $H_N$  of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H<sub>c</sub> of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated H<sub>c</sub> of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain  $H_N$  portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain  $H_N$  portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said second polypeptide is

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

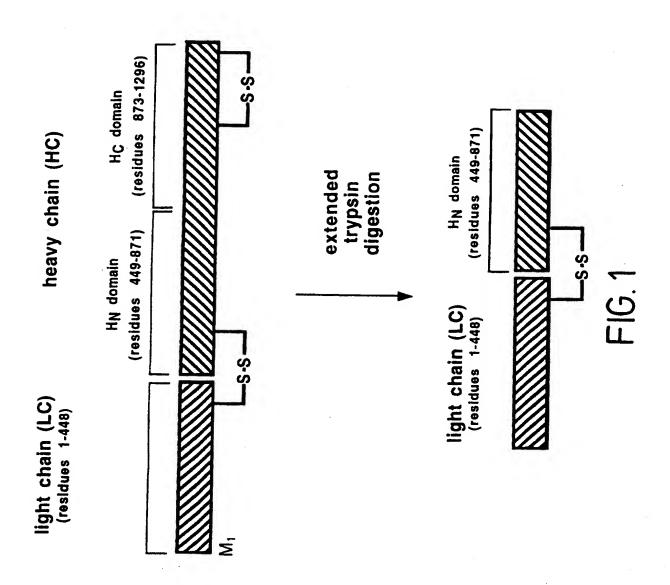
- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

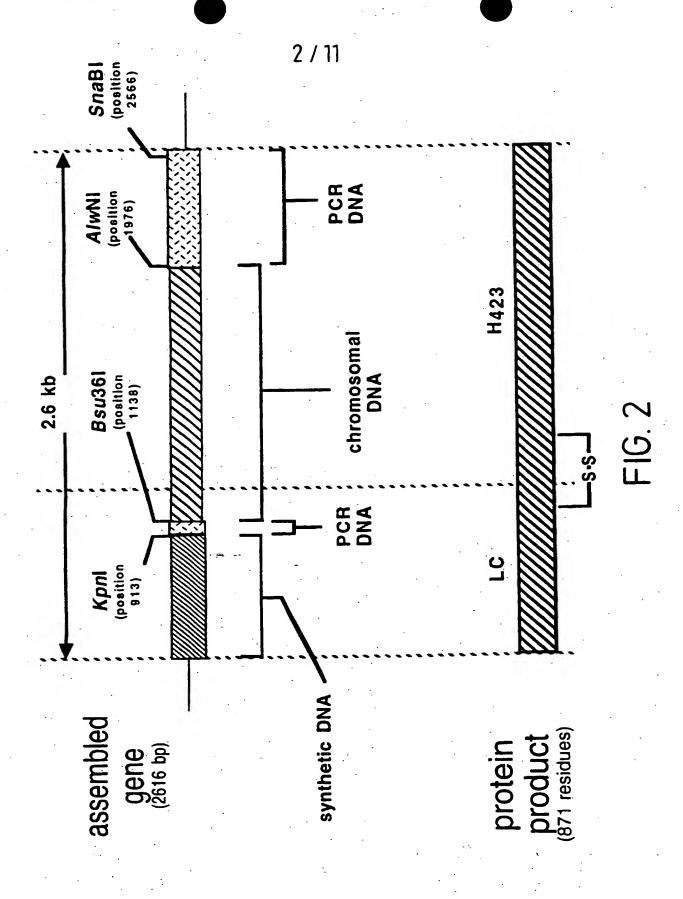
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

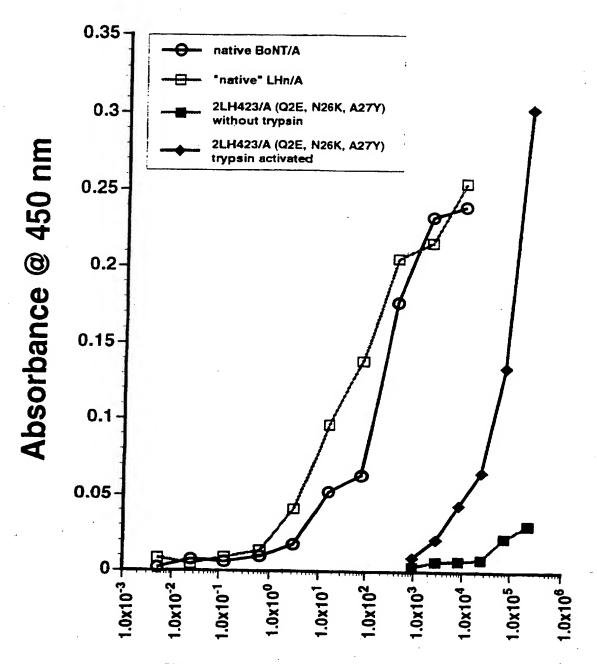
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain  $H_N$  domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain  $H_N$  domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.







Protein concentration (ng/ml)

FIG. 3

FNYKDPVNGVDIAYIKIPNAGOMOPV (SeqID.2)		FNYKDPVNGVDIAYIKIPKYGOMOPV (Seq1D4)	FNYKDPVNGVDIAYIKIPKYGQMQPV (Seq!D 6) +7	P N A G Q M Q P V	P N A G Q M Q P V
DPVNGVDIAYIKI	TRLQKLLEFELP	DPVNGVDIAYIKI	DPVNGVDIAYIKI	FNYKDPVNGVDIAYIKIPNAGQMQP	FNYKDPVNGVDIAYIKIPNAGQMQP
MQFVNKQFNYK 1	GSPGIHMTSTRLQ 1	E F V N K Q	E F V N K Q	M Q F V N K Q F N Y K	MPFVNKQFNYK
LH423/A 1	23LH423/A	K, A <sub>27</sub> Y)	2LH423/A G S N (Q2E, N26K, A27Y) 1 2 3	Native BoNT/A, C. botulinum 2169 Thompson et al.1990	Native BoNT/A, C. botulinum 62A

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.

F16. 4

1321/441 TCA TTA GAT AAA GGA TAC AAT AAG agc gct gat ggg GCA TTA AAT GAT TTA TGT ATC AAA S L D K G Y N K S A D G A L N D L C I K Eco47 III

F16. 5

1321/441

TCA TTA GAT AAA GGA TAC AAT AAG atc gaa ggt cgt tgc gat ggg GCA TTA AAT GAT TTA

S L D K G Y N K I E G R C D G A L N D L Factor Xa protease motif

CtxA14

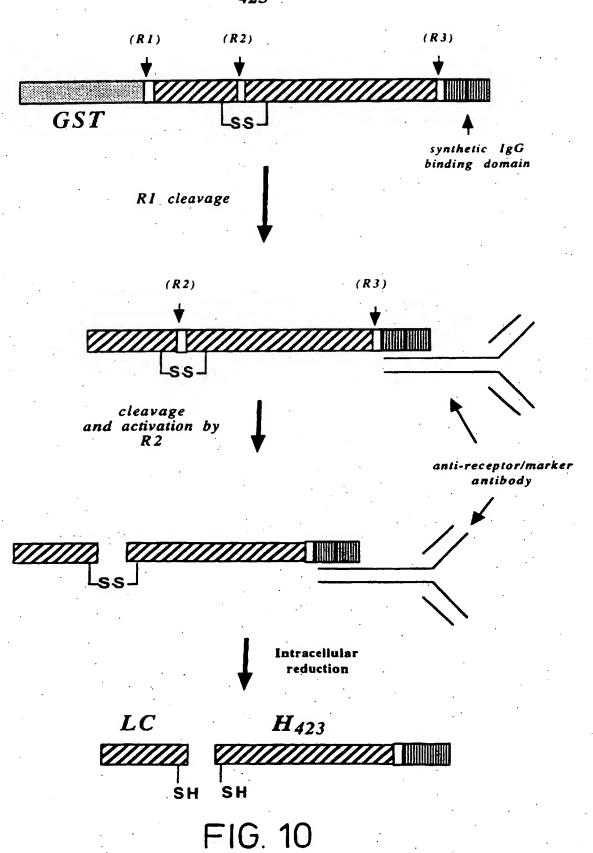
AGA R A S CAT AAT H N ACA T TCT S AAG K GAT D ATT ATT I GAT TAT Y TCT S GAA E 2617/873 TTT ACT F T 2677/893 TAT CAA Y Q ACA T TCA TCT S Eu TT A L TTA L ATA TGA stop AGA R 80 TTA L A O AGA R SAA E AAT N AAA K GAT D F5 > GAT D 

F16.8

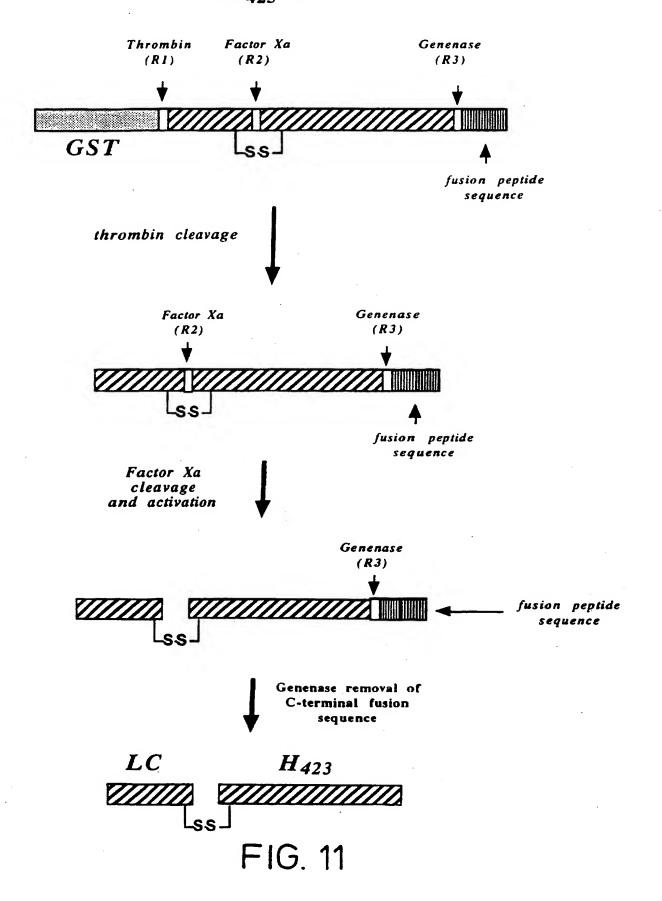
**₩** AGG R 767 C ACA T CAG 0 GAC D TAT Y ATG M egg g CCT P GAG E TCT S GCG A 161 C AAG K 6TG . AGG R CTG L TTC: C66 AGG R TAT CAG AGT S AGG R AGC S CTA 2617/873 TTT ACT F T 2677/893 GAT GCT D A 2737/913 GGC TCC G S 2797/933 TGT GAT C D 2857/953 GCT TAG GAA F 6TG V AGC S TAT Y TCT S CTG L L GGG G CGG R GCT A S CA TTA L ACA T GAG E T F AAG K TT A GCT. <u>)</u> 767 C AGA R 000 P 999 AAG K 76C C A S GAG 76C C AAC CCT 9 AAG K TTC F GAT D CTC L 7 6TG V GAT D ACG TAT Y 2587/863 1AC GTA (Y V L 2647/883 CCG GAG (P E E Z707/903 GGC TTT G F L G F Z767/923 GGT ATC G I L G I Z827/943 GCA CCC A P P P

A A A CGA GGC GCC AAT AAA K CTG L CAA Q GCA A GAA E AAC N AAC N GCT A وور وور TTC F AAA K GAA E TTA L GAA E CGA R TCA. AAA K GAA E E \_ \_ AAC N CAA Q AAG K AAC TTA AAC AAC N TTC F GAA E ATT I GAC D TTA L AAA K GCT A GAA E TAT Y GTA V AAC N AAC N AGC S AAC N 2617/873
TTT ACT GAA 1
F T E N
2677/893
GAT GAA GCC CAT TTA CCT N
H L P N
2797/933
CCA AGC CAA P
P S Q S
2857/953
AAA GTA GAC A
K V D N
2917/973
CCT AAC TTA A
P N L N
2977/993
CAA AGC GCT A
CAA AGC GCT A , GCG CCG A A P TTA CAT TT ACA T TTA L GAC D A F CAA O TCT S ATC I GAT D CCA P TTA L 900 A GAG E AAA K GAC D TTA L TAT Y TTA L GCT A GAT D ATC I AAA K AGA R T F AGT S GAG E GAT D S CAG Z= & & & o A GCG A TA L A A AAT N TAT Y A SCA CTA L AAC N ATC I TTC F GAT AGT S 2587/863
TAC GTA GAT
Y V D
2647/883
TCC CCG GGT
S P G
2707/903
GAA CAA CAA
E Q Q
2767/923
AAC GCC TTC
N A F
2827/943
GCT AAA AAG
A K
2887/963
CAA AAC GCG
Q N A
2947/983
TTC ATC CAA
F I Q
3007/1003
AAG CTA AAT
K I N

## 8/11 LH<sub>423</sub>/A



## $LH_{423}/A^{9/11}$



10/11

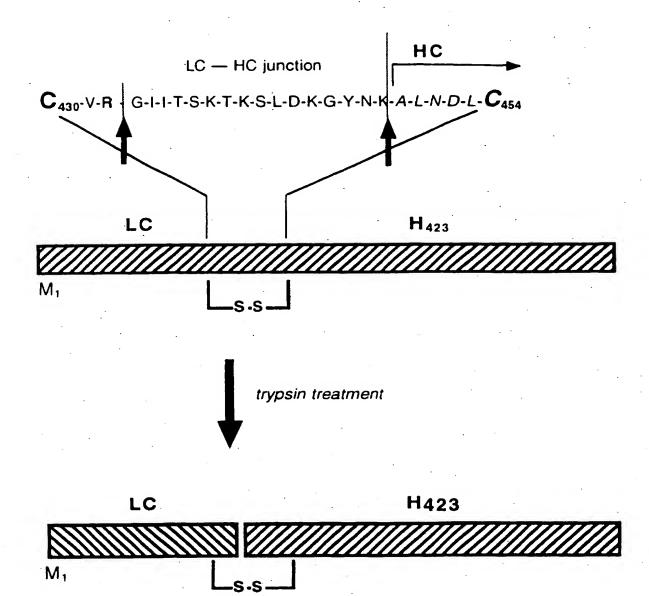
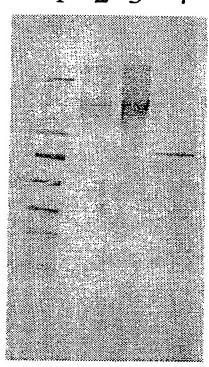


FIG. 12

11 / 11

Panel A. 1 2 3 4



Panel B.

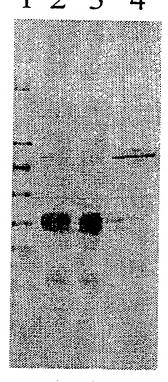


FIG. 13

### INTERNATIONAL SEARCH REPORT

pplication No Inten PCT) 97/02273

A. CLASSIFICATION OF SUBJECT MATTER
1PC 6 C12N15/31 C12N1/21 A61K39/08

C12P21/02

C07K14/33

A61K38/16

According to International Patent Classification (iPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) IPC .6 C12N C12P A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 12802 A (OPHIDIAN PHARM INC; WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document	1-52
X	KURAZONO H ET AL: "Minimal essential *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II	1-52
-	-/	
	*	
1		

X Further documents are listed in the continuation of box C.	X Patent family members are listed in annex.
*A* document defining the general state of the art which is not considered to be of particular relevance  *E* earlier document but published on or after the international filing date  *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  *O* document referring to an oral disclosure, use, exhibition or other means  *P* document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.  "8" document member of the same patent farmily
Date of the actual completion of the international search	Date of mailing of the international search report
9 December 1997	3 0. or 98
Name and mailing address of the ISA  European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Risswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016	Authorized officer Hillenbrand, G

### INTERNA DNAL SEARCH REPORT

Internal Application No PC17GB 97/02273

		PC17GB 97/02273		
.(Continua	ation) DOCUMENTS CONSIDERED TO BE RELEVANT			
ategory *	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.	
· ·	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document	·	1	
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document		1,26,35	
	·			
		•		
•				
			8	
	·			
	·	•		
			·	
	-			
	1			

## INTERNATIONAL SEARCH REPORT

PCT/P97/02273

•					PCT/ ST/GEETS				
Patent document cited in search report			Publication date	Patent family member(s)			Publication date		
WO	9612802	Α	02-05-96	US	5601823	A	11-02-97		
			•	US	5196193	Α	23-03-93		
•				AU	3968395	Α	15-05-96		
				EΡ	0796326	A	24-09-97		
	•			FI	971732	Α	23-06-97		
				NO	971868	Α	<b>24-06-97</b>		
				PL	320214	A ·	15-09-97		
				ZA	9508990	A	15-05-96		
				AU	6653894	Α	04-07-94		
			,	CA	2150935		23-06-94		
				EP	0671902		20-09-95		
			•	WO	9413264	Α	23-06-94		
				US	5466672		14-11-95		
•				US	5599539		04-02-97		
			•	AU	638786	В	08-07-93		
				AU	6895191		31-05-91		
	•			. EP	0498854		19-08-92		
				WO	9106306		16-05-91		
•	•			US	5443976		22-08-95		
			•	US	5340923	Α	23-08-94		

# THIS PAGE BLANK (USPTO)

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

#### **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

#### IMAGES ARE BEST AVAILABLE COPY.

OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

## THIS PAGE BLANK (USPTO)